

PRO GRADU -TUTKIELMA

Essi Syrjälä

**Lasten painoindeksien mallinnus ja painoindeksin käänne­pisteen
ajankohtaan yhteydessä olevien tekijöiden tutkiminen**

TAMPEREEN YLIOPISTO
Informaatiotieteiden yksikkö
Tilastotiede
Toukokuu 2016

Tampereen yliopisto

Informaatiotieteiden yksikkö

Syrjälä, Essi: Lasten painoindeksien mallinnus ja painoindeksin käännepisteen ajankohtaan yhteydessä olevien tekijöiden tutkiminen

Pro gradu -tutkielma, 57 s., 10 liites.

Tilastotiede

Toukokuu 2016

Tiivistelmä

Ihmisten lihavuus on huomattavasti lisääntynyt viime vuosikymmeninä maailmanlaajuisesti. Sen lisäksi, että jo yksistään lihavuus saattaa aiheuttaa yksilön elintason huononemista, se myös altistaa useille vakavillekin sairauksille, jotka huonontavat elintasoa ja myös lisäävät terveydenhuollon kustannuksia. Lihavuudesta eroon pääseminen on aina työlästä, mutta siihen takaisin palaaminen helppoa. Tämän vuoksi tulevaisuudessa olisi tärkeää oppia uusia keinoja ennaltaehkäistä lihavuutta jo varhaisessa vaiheessa elämää, varhaislapsuudesta lähtien. Lihavuuden kehittymisen vaiheita varhaislapsuudessa ja lapsuudessa on tutkittu jonkin verran, mutta osa mekanismeista on vielä tuntemattomia. Lihavuus on ainakin osittain perinnöllistä ja myös huonot elintavat saattavat siirtyä vanhemmilta lapselle. Vanhempien lihavuus ei kuitenkaan aina tarkoita, että lapsestakin tulisi lihava tai päinvastoin normaalipainoisten vanhempien lapsesta saattaa tulla lihava.

Lihavuuden määrittämiseen voidaan käyttää erilaisia mittareita, mutta ehkä eniten käytetty on painoindeksi (BMI), joka on paino (kg) jaettuna pituuden (m) neliöllä. Painoindeksin avulla voidaan määrittää ainakin aikuisten lihavuutta, mutta lapsen lihavuuden määrittelyyn se ei niin hyvin sovi. Sen avulla voidaan kuitenkin tutkia lapsuuden kehitysvaiheita. Lasten painoindeksin kehityksessä on havaittu olevan ainakin kaksi niin sanottua kriittistä kautta. Ensimmäinen niistä on painoindeksin varhaislapsuuden piikki (BMI peak) eli kohta, jossa lapsen painoindeksi on korkeimmillaan noin yhdeksän kuukauden iässä. Varhaislapsuuden piikin jälkeen painoindeksi lähtee laskuun. Tämän aiheuttaa rasvasolujen kasvun loppuminen sekä lapsen aktiivisuuden lisääntyminen. Toinen havaittu kriittinen kausi painoindeksin kehityksessä on painoindeksin käännepiste (adiposity rebound), jossa painoindeksi puolestaan on matalammillaan noin 5.5 vuoden iässä. Käännepisteen jälkeen painoindeksi lähtee nousuun rasvasolujen alkaessa jälleen kasvamaan ja lisääntymään.

Tässä tutkielmassa on tavoitteena mallintaa lasten painoindeksejä iän suhteen syntymän ja kymmenen ikävuoden välillä. Mallien avulla estimoidaan kullekin lapselle painoindeksin varhaislapsuuden piikki sekä painoindeksin käännepiste. Mallinnus tehdään kahdessa osassa, erikseen painoindeksin piikin sekä painoindeksin käännepisteen estimointia varten. Kahdestatoista aineiston lapsesta koostuvan testijoukon avulla tarkastellaan visuaalisesti mallinnuksen onnistumista ja valitaan kumpaankin tarkoitukseen sopiva mallinnustapa sekä painoindeksin piikin ja käännepis-

teen estimoinnin kriteerit. Lasten painoindeksien mallinnukseen syntymän ja kolmen ikävuoden välillä valitaan polynomiregressiomalli ja painoindeksin piikki määritetään olevan polynomiregressiomallin suurin estimaatti syntymän ja puolentoista ikävuoden välillä. Lasten painoindeksien mallinnukseen kahden ja kymmenen ikävuoden välillä valitaan kuutiotasoitettu splini ja painoindeksin käänneaste määritetään olevan splinikäyrän pienin estimaatti. Tutkittavien lasten estimoitu painoindeksin varhaislapsuuden piikki on keskimäärin noin kahdeksan kuukauden iässä, pojilla aikaisemmin kuin tytöillä. Lasten estimoitu painoindeksin käänneaste on keskimäärin 5.6 vuoden iässä, tytöillä hieman aikaisemmin kuin pojilla.

Tutkielman toisena tarkoituksena on tutkia painoindeksin käänneasteen ajankohtaan vaikuttavia tekijöitä. Aiemmissa tutkimuksissa painoindeksin käänneasteen aikaistuneen ajankohdan on osoitettu olevan yhteydessä nuoruuden sekä aikuisiän suurempaan painoindeksiin. Painoindeksin käänneasteen ajankohdan määräävät tekijät ovat vielä suurilta osin tuntemattomia. Aiemmassa kirjallisuudessa vanhempien painoindeksi sekä lapsen painoindeksin kehitys ennen käänneastetta on yhdistetty käänneasteen ajankohtaan. Tämän tutkielman tulokset tukevat näitä löydöksiä, mutta huomattavaa on, että tutkielman mikään malli ei selitä kovin suurta osaa lapsen painoindeksin käänneasteen ajankohdan vaihtelusta. Toisin sanoen tässä tutkielmassa tutkitut tekijät eivät ole painoindeksin käänneasteen määrääviä tekijöitä, vaikkakin osalla niistä on merkitsevää vaikutusta painoindeksin käänneasteen ajankohtaan.

Asiasanat lineaarinen regressioanalyysi, lineaarinen sekoitusregressiomalli, polynomiregressiomalli, kuutiotasoitettu splini, lihavuus

Sisältö

1	Johdanto	7
2	Tutkimuksen taustaa	9
2.1	Aiemmat tutkimukset ja menetelmät	9
2.1.1	Lapsen varhaislapsuuden kehityksen yhteys käännepesteeeseen	9
2.1.2	Ravinnon ja sosieekonomisten sekä vanhempin liittyvien te- kijöiden yhteys käännepesteeeseen	11
2.1.3	Yhteenveto	12
3	Tutkimusaineisto	14
3.1	Aineiston keräys	14
3.2	Aineiston kuvaus	14
3.2.1	Lasten syntymämittaukset	15
3.2.2	Äitien tiedot	17
3.2.3	Selittävien muuttujien väliset yhteydet	21
4	Tutkimuksessa käytetyt menetelmät	24
4.1	Tasoittava splini	24
4.1.1	Luonnollinen kuutiofasoitettu splini	25
4.1.2	Tasotusparametrin λ valinta	26
4.2	Lineaarinen regressiomalli	27
4.2.1	Hypoteesien testaus	29
4.2.2	Polynomiregressiomalli	30
4.3	Lineaarinen sekoitusregressiomalli	31
4.3.1	Akaiken informaatiokriteeri	32
5	Aineiston analyysi	34
5.1	Painoindeksikäyrien mallinnus	34
5.1.1	Painoindeksin varhaislapsuuden piikin mallinnus	34
5.1.2	Painoindeksin käännepesteen mallinnus	36
5.2	Lapsen painoindeksin käännepesteen ajankohtaan vaikuttavat tekijät	39
5.2.1	Yksinkertaiset lineaariset regressiomallit	39
5.2.2	Useamman selittäjän lineaariset regressiomallit	45
5.2.3	Lineaariset sekoitusregressiomallit	49
6	Johtopäätelmät	52
	Lähteet	55
A	Muuttujalista	58
B	Kyselylomakkeet	60

C	Mallien diagnostiikkaa	63
C.1	Malli 5.4	63
C.2	Malli 5.5	64
C.3	Malli 5.6	65
C.4	Malli 5.7	66
C.5	Sekoitusregressiomalli	67

1 Johdanto

Lihavuudesta on muodostunut maailmanlaajuisesti suuri kansanterveydellinen ongelma. Vuonna 2012 ylipainoisia ihmisiä oli maailmassa arviolta yli miljardi ja heistä noin kolmasosa oli lihavia. Lihavuus on vähintään tuplaantunut viimeisen 30 vuoden aikana. (Männistö, Laatikainen & Vartiainen (2012)). Myös juuri julkaistussa lähes koko maailman väestön kattavassa tutkimuksessa todettiin, että vuodesta 1975 vuoteen 2014 miesten painoindeksi on noussut keskimäärin 2.5 kg/m^2 ja naisten 2.3 kg/m^2 ((2016) (NCD-RisC)). Painoindeksi (BMI) on aikuisten ylipainoisuuden ja lihavuuden määrittelyssä yleisesti käytetty mittari. On olemassa muitakin tapoja määrittää ylipainoa ja lihavuutta, mutta tässä tutkielmassa on käytetty painoindeksiä. Painoindeksi lasketaan painon ja pituuden neliön suhteena $\text{paino}(\text{kg})/\text{pituus}(\text{m})^2$. Ylipaino ja lihavuus luokitellaan painoindeksin mukaan niin, että ylipainoiseksi luokitellaan, jos painoindeksi on $25 < 30 \text{ kg/m}^2$ ja lihavaksi, jos painoindeksi on 30 kg/m^2 tai suurempi. (Mustajoki (2015))

Suomalaisista miehistä kaksi kolmasosaa ja naisista puolet olivat ylipainoisia sekä kaikista suomalaisista viidesosa lihavia vuonna 2012. Ylipaino ja lihavuus ovat siis myös Suomessa sekä kansanterveydellisesti että -taloudellisesti merkittävä ongelma. Sen lisäksi, että ylipainoisuus ja lihavuus itsessään usein heikentävät elämänlaatua, ne myös altistavat useille muille sairauksille, kuten tyypin 2 diabetekselle, sydän- ja verisuonitauksille sekä tuki- ja liikuntaelinsairauksille. Tämä taas nostaa huomattavasti terveydenhuollon kuluja. (Männistö et al. (2012)). Ylipainon ja erityisesti lihavuuden lisääntyminen ovat tällä hetkellä ajankohtaisia asioita maailmanlaajuisesti. Tämän vuoksi olisikin tärkeää ymmärtää paremmin lihavuuden kehittymistä, jotta sen ennaltaehkäisy voitaisiin aloittaa jo varhaislapsuudessa.

Ihmisen painoindeksin kehitys voidaan jakaa kolmeen osaan. Ensimmäisen elinvuoden aikana painoindeksi kasvaa rasvasolujen kasvun myötä. Kun tämän kauden jälkeen rasvasolujen kasvu loppuu ja painoindeksi alkaa laskea lapsen kasvaessa pituutta ja alkaessa liikkumaan, on painoindeksikäyrästä havaittavissa painoindeksin varhaislapsuuden piikki eli infant BMI peak. Tämä on kohta, missä painoindeksi on korkeimmillaan ensimmäisen elinvuoden aikana. Se esiintyy keskimäärin noin yhdeksän kuukauden iässä. Tämän jälkeen painoindeksi kääntyy laskuun. Lasku jatkuu keskimäärin noin viiden-kuuden vuoden ikään asti, jonka jälkeen alkaa kolmas kausi. Tällöin rasvasolut alkavat taas kasvaa ja niiden määrä lisääntyy, jolloin painoindeksi alkaa kasvaa, kunnes se saavuttaa lopullisen tasonsa. Kolmannen kauden alkua, jolloin painoindeksi on matalimmillaan, kutsutaan painoindeksin käännepisteeksi eli adiposity reboundiksi (AR). (Rolland-Cachera, Deheeger, Bellisle, Bellisle, Sempe, Guilloud-Bataille & Patois (1984))

Painoindeksin käännepisteen ajankohdalla on todettu olevan yhteys nuoruusiän korkeampaan painoindeksiin sekä tämän seurauksena nuoruuden ja aikuisiän ylipainoisuuteen siten, että aikaisempi painoindeksin käännepiste altistaa suurempaan ylipainoisuuden ja lihavuuden riskiin (Rolland-Cachera et al. (1984), Whitaker, Pepe, Wright, Seidel & Dietz (1998), Eriksson, Forsén, Tuomilehto, Osmond & Barker

(2003), Taylor, Grant, Goulding & Williams (2005), Hughes, Sherriff, Ness & Reilly (2014)). On myös todettu, että ikä, jolloin käännepiste esiintyy, on mahdollisesti aikaistunut vuosien saatossa (Vignerová, Humeníkova, Brabec, Riedlová & Bláha (2007), Nummi, Hakanen, Lipiäinen, Harjunmaa, Salo, Saha & Vuorela (2014)). Tämä havainto olisikin samassa linjassa sen tiedon kanssa, että ylipainoisuus ja lihavuus ovat viime vuosikymmeninä huomattavasti lisääntyneet. On kuitenkin vaikea sanoa varmaksi, onko käännepisteen ajankohdassa tapahtunut muutosta, sillä käännepiste on määritetty eri tavoin eri tutkimuksissa (Taylor et al. (2005)).

Tässä tutkielmassa mallinnetaan suomalaisten lasten painoindeksin kehitystä syntymästä kymmenen vuoden ikään saakka. Mallinnus on tehty kahdessa osassa kahdella eri tavalla, käyttäen neljännen asteen polynomiregressiomalleja sekä käyttäen kuutiotasoitettuja splineja. Mallinnetuilta painoindeksikäyryltä on määritetty painoindeksin varhaislapsuuden piikki sekä painoindeksin käännepiste. Painoindeksin piikki on määritetty olevan polynomifunktion maksimiarvo eli polynomikäyrän korkein kohta ja painoindeksin käännepiste on määritetty olevan splinikäyrän minimikohta. Lisäksi tässä tutkielmassa tutkitaan äidin painoindeksin, painon, pituuden ja diabeteksen sekä lapsen syntymämittausten (pituus, paino, painoindeksi ja päänympärys) ja painoindeksin piikin ajankohdan ja piikin aikaisen painondeksin yhteyttä painoindeksin käännepisteen ajankohtaan. Nämä tekijät on valittu aikaisemman kirjallisuuden pohjalta. Koska aihetta on tutkittu vielä melko vähän, on tutkittaviksi tekijöiksi valittu myös sellaisia, joilla on jo osoitettu olevan yhteys käännepisteen ajankohtaan. Malleihin sisällytetään myös muita saatavilla olevia muuttujia ja muutkin mahdolliset käännepisteeseen yhteydessä olevat tekijät raportoidaan, jos sellaisia löytyy. Aineistoa on analysoitu käyttäen lineaarisia sekoitusregressiomalleja, joihin sisältyy myös tavallinen lineaarinen regressiomalli.

2 Tutkimuksen taustaa

Painoindeksin käännepesti saattaa olla yksi kriittisistä kausista ylipainon ja lihavuuden kehittymisen kannalta. Se, että suurimmalla osalla lihavista aikuisista painoindeksin käännepestin ajankohta on ollut aikaistunut, antaa viitteitä siitä, että lihavuuden taustatekijät saattavat löytyä jo elämän varhaisista vaiheista. (Rolland-Cachera, Deheeger, Maillot & Bellisle (2006)). Painoindeksin käännepestin yhteyttä aikuisiän ylipainoisuuteen ja lihavuuteen on tutkittu melko paljon useissa maissa ja vahva yhteys on selkeästi olemassa. Käännepestin ajankohtaan vaikuttavia tekijöitä puolestaan on tutkittu vielä melko vähän ja niitä ei juurikaan tunneta. Aihetta on tarpeellista tutkia lisää, jotta voidaan oppia ymmärtämään paremmin lihavuuden kehittymistä ja sen eri vaiheita jo varhaislapsuudesta lähtien. Aikaistuneelle käännepestille altistavien tekijöiden tunnistaminen olisi suuri etu lihavuuden riskiryhmien määrittämisen kannalta ja näin voitaisiin kohdistaa lihavuuden ennaltaehkäisy paremmin sitä tarvitseville.

Toinen mahdollisista kriittisistä kausista on painoindeksin varhaislapsuuden piikki. Siihen liittyviä tutkimuksia ei ole vielä tehty kovin paljoa, mutta senkin tiedetään olevan yhteydessä ainakin lapsuuden ja nuoruuden ylipainoon ja lihavuuteen (Silverwood, De Stavola, Cole & Leon (2009), Mook-Kanamori, Durmus, Sovio, Hofman, Raat, Steegers, Järvelin & Jaddoe (2011), Hof, Vrijkotte, de Hoog, van Eijdsen & Zwinderman (2013)). Myös painoindeksin piikkiin liittyviä tutkimuksia kaivattaisiin lisää. Tässä tutkielmassa painoindeksin piikin ajankohtaa sekä piikin aikaista painoindeksiä on käytetty selittävinä muuttujina painoindeksin käännepestille. Sekä painoindeksin varhaislapsuuden piikki että painoindeksin käännepesti on määritetty aikaisemmassa kirjallisuudessa monin eri tavoin.

Vanhempien lihavuuden tiedetään altistavan lapsen selvästi suurempaan lihavuuden riskiin. Whitaker et al. (1998) totesivat tutkimuksessaan, että aikaistuneen painoindeksin käännepestin yhteys aikuisiän lihavuuteen tulee esille riippumatta vanhempien lihavuudesta. Tämä tarkoittaa sitä, että lapsen käännepestin ollessa aikaistunut, on lapsella suurempi lihavuuden riski, vaikka vanhemmat olisivatkin normaali-painoisia. Lisäksi on huomattu, että lapsuuden painoindeksi ennen käännepestettä ennustaa huonosti aikuisuuden lihavuutta (Rolland-Cachera et al. (2006)). Tämän vuoksi onkin mielenkiintoista ja tärkeää tutkia painoindeksin kehityksen varhaisia vaiheita ja siihen liittyviä kriittisiä kohtia, joihin painoindeksin käännepesti lukeutuu, ja oppia ymmärtämään niiden syitä ja seurauksia.

2.1 Aiemmat tutkimukset ja menetelmät

2.1.1 Lapsen varhaislapsuuden kehityksen yhteys käännepesteseen

Syntymäpainon sekä varhaislapsuuden painoindeksin kehitysvaiheiden yhteydestä käännepesteseen on aiemmassa kirjallisuudessa hieman ristiriitaista tietoa. Eriksen, Kajantie, Lampl, Osmond & Barker (2014) eivät löytäneet tutkimuksessaan yh-

teyttä syntymäpainon ja käännepesteen ajankohdan välille. Wen, Kleinman, Gillman, Rifas-Shiman & Taveras (2012) puolestaan havaitsivat tutkimuksessaan, että korkeampaa syntymäpainoa seurasi aikaisempi käännepesti sekä korkeampi painoindeksi käännepestiessä. He mallinsivat painoindexejä sekavaikutusmallien ja murtopolynomifunktioiden avulla. Myös Koyama, Kariya, Ichikawa, Shimura, Sairenchi & Arisaka (2013) löysivät tutkimuksessaan yhteyden lapsen korkean syntymäpainon (yli 3500g) ja aikaistuneen käännepesteen välille. Alhaisempien syntymäpainojen ja käännepesteen välille yhteyttä ei löytynyt.

Eriksson et al. (2003) löysivät tutkimuksessaan yhteyden lapsen alhaisemman yksivuotispainon ja -painoindexin sekä ensimmäisen elinvuoden hitaamman painonnousun ja aikaistuneen painoindexin käännepesteen välille. Myöhemmässä tutkimuksessaan Eriksson et al. (2014) testasivat hypoteesia, jonka mukaan sikiön kasvu, joka määriteltiin lapsen kokona syntyessä (pituus, paino, päänympäryys, pää poikittainen halkaisija) olisi yhteydessä painoindexin käännepesteen ajankohtaan. Molemmissa tutkimuksissa painoindexin käännepesti oli määritelty olevan lapsen pienin painoindeksi yhden ja 12 ikävuoden välillä. Syntymämittauksista ainoastaan lapsen pienempi päänympäryys sekä pienempi pää poikittainen halkaisija olivat yhteydessä aikaisempaan käännepesteen ajankohtaan. Yhteys varhaislapsuuden matalan painoindexin ja aikaistuneen käännepesteen välille tuli esille kuuden kuukauden iästä lähtien neljän vuoden ikään asti. Neljän ikävuoden jälkeen yhteys kääntyi päinvastaiseksi, tämän iän jälkeinen korkea painoindeksi oli yhteydessä aikaistuneeseen käännepestieseen. Myös Johnson, Soloway, Erickson, Choh, Lee, Chumlea, Siervogel, Czerwinski, Towne & Demerath (2012) havaitsivat tutkimuksessaan, että matalampaa painoindexiä viiden vuoden ikään asti seuraa usein nuoruusiän nopeampi painoindexinnousu.

Rolland-Cachera et al. (1984) puolestaan havaitsivat tutkimuksessaan, että korkeampaa rasvamäärää yhden vuoden iässä seurasi aikainen painoindexin käännepesti. Käännepesti oli määritetty olevaksi pienin painoindexin arvo ennen painoindexin kasvun alkamista. Tämä tutkimustulos on hieman ristiriidassa muiden, edellä mainittujen, tutkimustulosten kanssa.

Myös painoindexin varhaislapsuuden piikin, joka on keskimäärin noin yhdeksän kuukauden iässä, yhteyttä käännepesteen painoindexiin sekä lapsuuden ja nuoruuden painoindexiin on tutkittu jonkin verran. Silverwood et al. (2009) tutkivat painoindexin varhaislapsuuden piikin yhteyttä myöhäisemmän lapsuuden painoindexiin. Myöhäisemmän lapsuuden painoindexin mittauskohta vaihteli viiden ja kolmentoista ikävuoden välillä. Keskimäärin mittaus oli tehty kymmenen vuoden iässä. Heidän tutkimuksessaan myöhäisempi painoindexin piikin ikä sekä suurempi painoindeksi piikin aikaan olivat yhteydessä suurempaan painoindexiin myöhäisemmässä lapsuudessa. Tämän seurauksena voisi ajatella, että nämä tekijät saattaisivat olla yhteydessä myös aikaistuneeseen käännepesteen ajankohtaan, sillä se puolestaan on yhteydessä suurempaan painoindexiin myöhäisemmässä lapsuudessa ja nuoruudessa. He mallinsivat painoindexikäyrät käyttäen sakotettuja splineja satunnaisilla kertoimilla ja painoindexin piikki oli splinikäyrän maksimikohta.

Wen et al. (2012) puolestaan löysivät vahvan positiivisen korrelaation (0.76) varhaislapsuuden piikin painoindexin ja käännepesteen painoindexin välille. He käyt-

tivät sekavaikutusmalleja ja murtopolynomeja mallintamaan painoindeksijä. Painoindeksin piikki ja käännepestite olivat käyrien käännepestiteitä. Mook-Kanamori et al. (2011) tekivät samansuuntaisen löydöksen todetessaan, että suuri paino varhaislapsuuden piikin aikaan suurentaa huomattavasti riskiä ylipainoisuuteen tai lihavuuteen neljän vuoden iässä. He käyttivät logaritmoitujen painoindeksien mallinnukseen kuutiollisia sekavaikutusmalleja. Hof et al. (2013) tutkivat yhteyttä painoindeksin varhaislapsuuden piikin ja 5-6-v kehonkoostumusmittausten välillä. Lapsien painoindeksin piikit poimittiin sekavaikutusmalleista erikseen tytöille ja pojille. Tutkimuksessa myöhäisempi painoindeksin piikki sekä korkeampi painoindeksi piikin aikaan johtivat suurempaan kehonkoostumusmittaus tuloksiin 5-6 vuoden iässä. Näiden tutkimusten perusteella näyttäisi vahvasti siltä, että piikin aikainen suuri painoindeksi ennakoii suurta painoindeksiä myös käännepestiteessä.

Toisaalta taas Dorosty, M. Emmett, Cowin & Reilly (2000) eivät löytäneet minikäänlaista yhteyttä varhaislapsuuden painoindeksin (alle 4-v.) ja käännepestiteen ajankohdan välille. He kuitenkin löysivät viitteitä siitä, että ainakin tytöillä, joilla käännepestiteen ajankohta oli erittäin aikainen, olisi muita korkeampi painoindeksi juuri ennen käännepestitettä. Yhden tutkimuksen perusteella myös suurempi painoindeksi neljän ikävuoden jälkeen oli yhteydessä aikaisempaan käännepestiteeseen (Eriksson et al. (2014)). Näiden perusteella koska painoindeksin piikin suuremman BMI:n on osoitettu olevan yhteydessä suurempaan painoindeksiin ikävuosina, jolloin painoindeksin käännepestite suurinpiirtein esiintyy (Wen et al. (2012), Mook-Kanamori et al. (2011) ja Hof et al. (2013)), voisi painoindeksin piikin suuremman BMI:n ajatella olevan yhteydessä myös aikaistuneeseen käännepestiteeseen. Kuitenkin muutamien tutkimusten perusteella pieni painoindeksi varhaislapsuudessa ennakoisi aikaisempaa käännepestitettä tai nuoruuden nopeampaa painonnousua (Eriksson et al. (2003), Eriksson et al. (2014) ja Johnson et al. (2012)). Aikaistunutta käännepestitettä edeltävistä kasvumalleista on siis olemassa tietoa, mutta tutkimustulokset ovat osittain ristiriidassa keskenään.

2.1.2 Ravinnon ja sosieekonomisten sekä vanhempiin liittyvien tekijöiden yhteys käännepestiteeseen

Rolland-Cachera, Deheeger, Akrouit & Bellisle (1995) löysivät tutkimuksessaan yhteyden varhaisen proteiininsaannin ja painoindeksin käännepestiteen ajankohdan välille. Mitä korkeampi oli proteiininsaanti kahden vuoden iässä, sitä aikaisempi oli painoindeksin käännepestiteen ajankohta. Painoindeksin käännepestite oli määritetty olevaksi pienin painoindeksin arvo ennen painoindeksin kasvun alkamista. Seurantatutkimus koostui 112 ranskalaisesta lapsesta eli otoskoko oli melko pieni. Tämän ranskalaisen tutkimuksen pohjalta Dorosty et al. (2000) testasivat hypoteesia, jonka mukaan korkea proteiininsaanti puolentoista vuoden iässä aikaistaisi painoindeksin käännepestiteen ajankohtaa. Lisäksi he tutkivat muun varhaisen ravinnonsaannin sekä vanhempien painoindeksin, sosieekonomisen aseman ja koulutuksen yhteyttä lapsen painoindeksin käännepestiteen ajankohtaan. Tutkimuksessa painoindeksin käännepestiteen ajankohta määritettiin olevan joko erittäin aikainen, aikainen tai keskiverito perustuen visuaaliseen määrittelyyn lapsen painoindeksin kehityksen perusteel-

la. Toisin kuin edellä mainitussa tutkimuksessa, tässä tutkijat eivät löytäneet yhteyttä varhaisen proteiininsaannin ja painoindeksin käännepisteen ajankohdan välille. Kummassakaan näistä tutkimuksista myöskään varhaislapsuuden muun ravinnonsaannin ei todettu olevan yhteydessä käännepisteen ajankohtaan. Sen sijaan yhteys vanhempien suuremman painoindeksin ja aikaistuneen käännepisteen välille löytyi. Lisäksi lapsilla, joiden vanhemmista ainakin toinen oli lihava ($BMI \geq 30$), oli merkittävästi suurempi riski aikaiseen tai erittäin aikaiseen käännepisteeseen. Myös Whitaker et al. (1998) totesivat jo aiemmin tutkimuksessaan vanhempien suuremman painoindeksin olevan yhteydessä aikaistuneeseen painoindeksin käännepisteen ajankohtaan. He mallinsivat lasten logaritmoituja painoindeksejä kolmannen asteen polynomifunktioilla ja määrittivät käännepisteen olevan sovitetun käyrän minimipiste. Käännepisteet luokiteltiin kolmeen ryhmään: aikainen, keskiverto ja myöhäinen.

Eriksson et al. (2014) eivät löytäneet tutkimuksessaan lapsen elinympäristöön liittyvien tekijöiden ja painoindeksin käännepisteen välille yhteyttä. Äidin loppuraskauden korkea painoindeksi sekä korkea pituus olivat yhteydessä aikaistuneeseen käännepisteeseen. Äidin aikaisempien synnytysten suurempi lukumäärä puolestaan oli yhteydessä myöhäisempään painoindeksin käännepisteeseen. Koyama et al. (2013) löysivät tutkimuksessaan myös myöhäisempään käännepisteeseen yhteydessä olevia tekijöitä, joihin lukeutui aamupalan syöminen joka aamu, napostelun välttäminen, ei-lihava isä, ensimmäinen lapsi ja tarhassa käyminen. Näistä ainakin ensimmäinen lapsi on hieman ristiriidassa sen kanssa, että Eriksson et al. (2014) löysivät yhteyden äidin aikaisempien synnytysten suuremman lukumäärän ja myöhäisemmän käännepisteen välille. Koyama et al. (2013) määrittävän käännepisteen olevan lapsen matalin painoindeksi neljän kuukauden ja 12 ikävuoden välillä ja jakoivat tämän jälkeen käännepisteen ajankohdat kahteen ryhmään (aikaisempi tai myöhäisempi käännepiste). Jos painoindeksi kolmen vuoden iässä oli suurempi kuin 1.5-vuotiaana, tämä määriteltiin aikaisemmaksi käännepisteeksi ja vastaavasti jos 3-v painoindeksi ei ollut suurempi kuin 1.5-v painoindeksi, kyseessä oli myöhäisempi käännepiste. Kyseessä on hiukan erikoinen tapa luokitella käännepistettä, kun tiedetään, että käännepiste esiintyy keskimäärin 5-6-vuoden iässä. Ongelmana on, että myöhäisempään käännepisteen ryhmään saattaa kuulua myös lapsia, joilla käännepiste on ollut keskimääräistä aikaisempi tai täysin keskiverto.

2.1.3 Yhteenveto

Painoindeksejä on aiemmassa kirjallisuudessa mallinnettu iän suhteen monin eri tavoin ja tämän myötä painoindeksin piikki sekä käännepiste on myös määritetty usealla eri tavalla. Painoindeksin piikki on useassa tutkimuksessa estimoitu sekavaikutusmallejen sekä polynomien avulla. Painoindeksin käännepistettä ei ole joissakin tutkimuksissa estimoitu ollenkaan, vaan on poimittu vain painoindeksin pienin arvo tietyltä väliltä. Toisissa tutkimuksissa taas käännepiste on määritetty visuaalisesti painoindeksikäyrästä. Nummi et al. (2014) mallinsivat painoindeksejä iän suhteen kuutiotasoitetuilla splineilla, joita käytetään myös tässä tutkielmassa. Heidän tutkimuksessaan mittapisteitä jokaista lasta kohden oli vain seitsemän. Tässä tutkielmas-

sa niitä on paljon enemmän. Lisäksi joissakin tutkimuksissa käännepisteen ajankoh-
taa on käsitelty luokiteltuna muuttujana. Tässä tutkielmassa käännepisteen ajankoh-
ta on jatkuva muuttuja, minkä ansiosta aineistosta on mahdollista saada tarkempia
tuloksia.

Tähän mennessä tehtyjen tutkimusten perusteella on löydetty joitakin käänne-
pisteen ajankohtaan yhteydessä olevia tekijöitä. Vanhempien suurempi painoindeksi
näyttäisi selkeästi olevan yhteydessä lapsen aikaistuneeseen painoindeksin käänne-
pisteeseen. Yhden tutkimuksen perusteella pieni pää syntyessä olisi yhteydessä ai-
kaistuneeseen käännepisteen ajankohtaan. Tässä tutkimuksessa myös äidin korkean
pituuden ja aikaisen käännepisteen välille löytyi yhteys. Kahdessa tutkimuksessa oli
havaittu yhteys korkean syntymäpainon ja aikaistuneen käännepisteen välille. Myös
lapsen varhaislapsuuden koon ja käännepisteen välisiä yhteyksiä oli havaittu, mutta
ne olivat hieman ristiriidassa keskenään. Oli esitetty myös, että suurempi painoin-
deksi juuri ennen käännepistettä olisi yhteydessä aikaisempaan käännepisteeseen.
Koska suurempi fyysinen aktiivisuus puolestaan on yhteydessä matalampaan kehon
rasvamäärään painoindeksin käännepisteen aikoihin, tulevaisuudessa olisi perustel-
tua tutkia, voitaisiinko aktiivisuutta lisäämällä viivästyttää käännepisteen ilmenty-
mistä (Taylor et al. (2005)).

3 Tutkimusaineisto

3.1 Aineiston keräys

Painoindeksikäyrien mallinnukseen käytetty aineisto koostuu tamperelaisista vuonna 2003 ja 2004 syntyneistä lapsista. Se on kerätty kaikille lapsille säännöllisesti tehdyissä terveystarkastuksissa, joissa seurataan lapsen kasvua ja kehitystä. Jokaisella käynnillä on mitattu lapsen pituus ja paino. Syntymämittaukset on tehty synnytysairaalassa ja niihin sisältyy pituuden ja painon lisäksi lapsen päänympärys. Lisäksi käytössä on aineistoja, jotka sisältävät lasten vanhemmista kerättyjä tietoja. Aineistoista löytyy taustatietoja vanhemmista sekä neuvolassa raskauden aikana ja synnytyksen jälkitarkastuskäynnillä kerättyjä tietoja liittyen raskauteen, synnytykseen ja synnytyksen jälkeiseen aikaan. Tietojen keräämiseen on käytetty neuvolassa täytettäviä lomakkeita. Lomakkeet löytyvät liitteestä B.

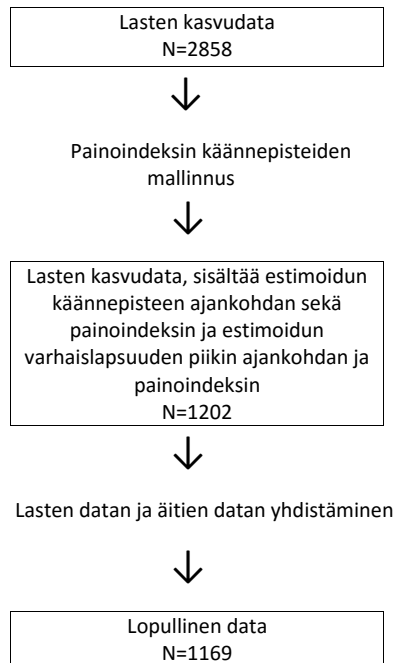
Suomessa terveystarkastukset on suunniteltu tehtäväksi jokaiselle lapselle tietynä ikäpisteinä. Nämä suunnitellut ikäpisteet ovat: 1-2 viikkoa, 3-4 viikkoa, 6-8 viikkoa, 2 kuukautta, 3 kuukautta, 4 kuukautta, 5 kuukautta, 6 kuukautta, 8 kuukautta, 10 kuukautta, 12 kuukautta, 18 kuukautta, 24 kuukautta, 36 kuukautta ja 48 kuukautta. Tämän jälkeen terveystarkastuksia tehdään noin vuoden välein. Lisäksi joitakin lapsia on seurattu tiuhempaan, jos heidän terveydessään tai kehityksessään on ollut jonkinlaisia ongelmia.

Lasten painot on mitattu elektronisella vaa'alla lapsen ollessa ilman vaatteita tai hyvin kevyessä vaatekannassa. Pituusmittaukset on tehty kaksivuotiaaksi asti tähän tarkoitukseen suunnitellulla mittauspöydällä ja tästä eteenpäin seisoalteen seiniä vasten mittanauhan avulla. (Harjunmaa (2009)).

3.2 Aineiston kuvaus

Lapsiin liittyvä alkuperäinen aineisto sisältää 2858 lapsen säännölliset pituus- ja painomittaukset sekä lapsen iän kullakin mittauskerralla. Pienin määrä mittauskertoja yhtä lasta kohden on yksi ja suurin määrä 53 mittauskertaa. Aineistossa on paljon lapsia, joilta löytyy vain muutamia mittauskertoja. Pieni mittausmäärä johtuu todennäköisesti siitä, että lapsi on muuttanut pois Tampereelta tai muuttanut vasta myöhemmin Tampereelle. Keskimäärin mittauskertoja on noin 18 yhtä lasta kohden. Painoindeksin käännepestettä mallinnettaessa lapselta täytyy löytyä tarpeeksi mittapisteitä, jotta mallinnus on järkevää tehdä. Lisäksi joillakin lapsilla painoindeksin käännepestettä ei ole selvästi nähtävissä eikä sitä näin ollen pystytä määrittämään. Näiden tekijöiden vuoksi painoindeksin käännepesteidien mallinnuksen seurauksena aineisto pieneni huomattavasti, ja kun lasten dataan vielä yhdistettiin äiteihin liittyvä data, lopullinen aineisto sisältää 1169 lapseen liittyvän datan. Kuviossa 3.1 on esitetty lopullisen aineiston kokoamisen vaiheet.

Seuraavassa alaosiossa on esitelty aineiston muuttujia. Lista ja kuvaus aineiston



Kuvio 3.1. Lopullisen aineiston kokoamisen vaiheet.

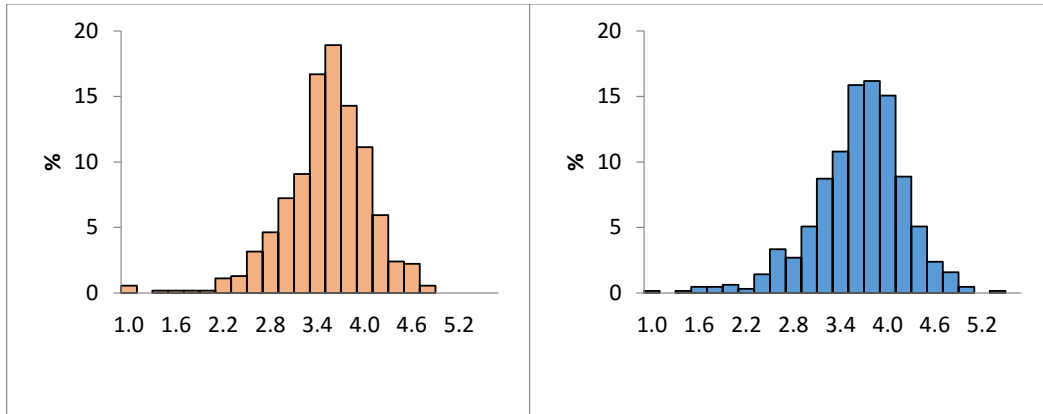
kaikista muuttujista löytyy liitteestä A.

3.2.1 Lasten syntymämittaukset

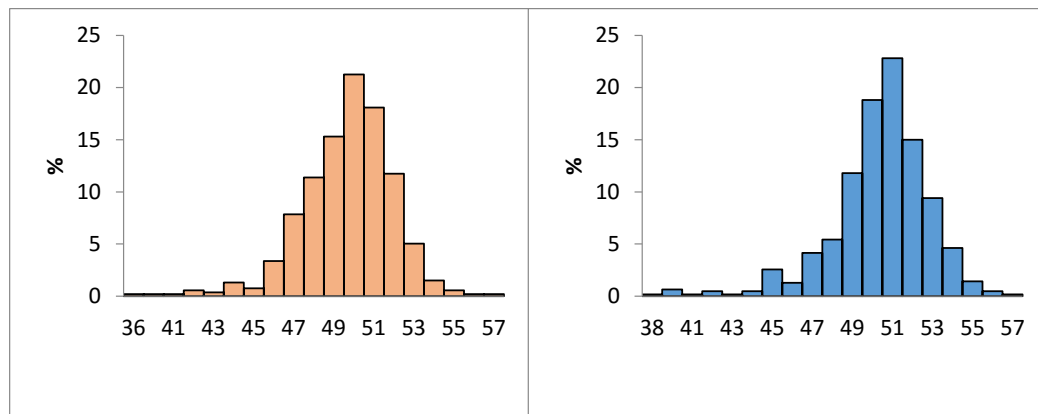
Aineiston lapsista 54% on poikia ja 46% tyttöjä. Lasten keskimääräinen syntymäpaino oli 3.5 kg (tytöt 3.4 kg ja pojat 3.6 kg) ja syntymäpituus 50.1 cm (tytöt 49.7 cm ja pojat 50.4 cm). Pojat olivat siis keskimäärin pidempiä ja painavampia kuin tytöt ($p < 0.001$). Kuviossa 3.2 on esitetty erikseen tyttöjen ja poikien syntymäpainojen jakaumat ja kuviossa 3.3 syntymäpituuksien jakaumat. Nähdään, että kaikki jakaumat painottuvat hieman oikealle, mutta ovat lähellä normaalijakaumaa. Tyttöjen ja poikien jakaumat eroavat jonkin verran toisistaan, poikien jakauma painottuu enemmän suurempiin arvoihin sekä syntymäpainon että pituuden suhteen, niin kuin oli odotettavissakin.

Lasten keskimääräinen syntymäpainoindeksi oli 13.9 (tytöt 13.8 kg/m² ja pojat 13.9 kg/m², ei tilastollisesti merkitsevää eroa). Pienin painoindeksi oli 6.9 ja suurin 22.8 eli syntymäpainoindekseissä oli paljon hajontaa. Tyttöjen ja poikien syntymäpainoindeksit olivat normaalisti jakautuneet ja jakaumat olivat hyvin samankaltaisia keskenään.

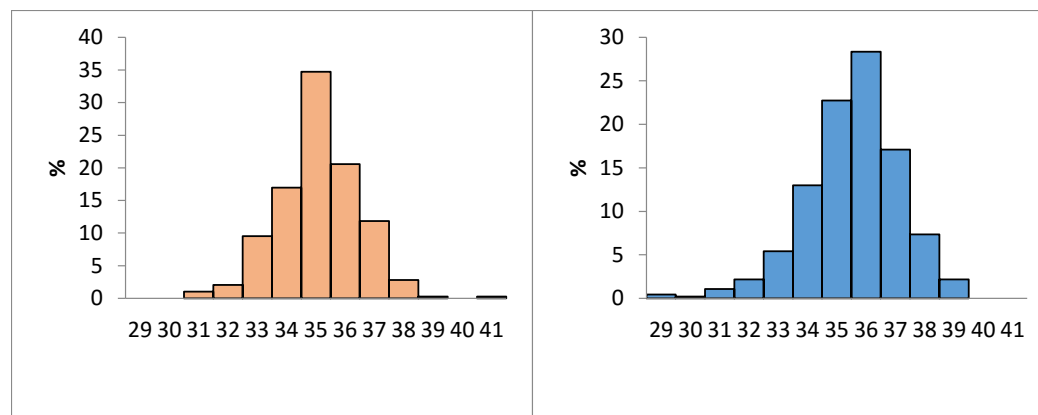
Lasten keskimääräinen päänympärysy syntyessä oli 35.1 cm (tytöt 34.9 cm ja pojat 35.4 cm, $p < 0.001$). Kuviossa 3.4 on esitetty erikseen tyttöjen ja poikien päänympäryksien jakaumat. Nähdään, että jakaumat ovat normaaleja ja eroavat jonkin verran toisistaan. Tyttöjen jakaumassa on selvä huippu noin 35 cm kohdalla, kun taas poikien jakauma on hieman tasaisempi. Syntymämittauksista paino löytyy aineiston kaikille lapsille, pituus ja tämän vuoksi myös painoindeksi puuttuu kuudelta lapselta



Kuvio 3.2. Vasemmalla tyttöjen ja oikealla poikien syntymäpainon (kg) jakauma.



Kuvio 3.3. Vasemmalla tyttöjen ja oikealla poikien syntymäpituuden (cm) jakauma.

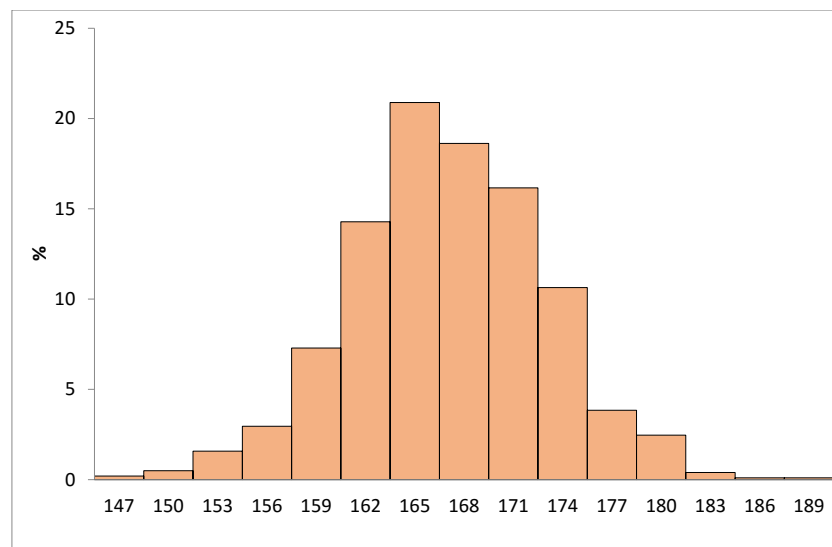


Kuvio 3.4. Vasemmalla tyttöjen ja oikealla poikien syntymäpään ympäryksen (cm) jakauma.

ja päänympärys puuttuu jopa 318 lapselta.

3.2.2 Äitien tiedot

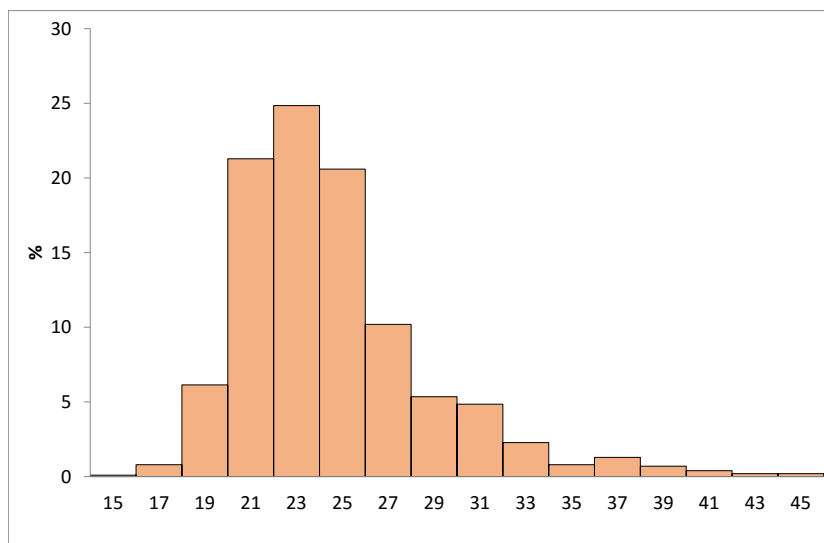
Äitien keskimääräinen paino ennen raskautta oli 65 kg, minimin ollessa 39 kg ja maksimin 129 kg. Painon mediaani oli 63 kg eli jakauma on hiukan oikealle vinoutunut. Äitien keskimääräinen pituus oli 166 cm, minimin ollessa 145 cm ja maksimin 187 cm. Äitien pituuden jakauma on esitetty tasavälisesti 5 cm:n välein luokiteltuna kuviossa 3.5. Kuviosta nähdään, että pituus noudattaa hyvin normaalijakaumaa. Tieto painosta ennen raskautta oli saatavilla 1011:lle äidille ja tieto pituudesta 1015:lle äidille. Näistä 1010:lle äidille pystyttiin määrittämään painoindeksi.



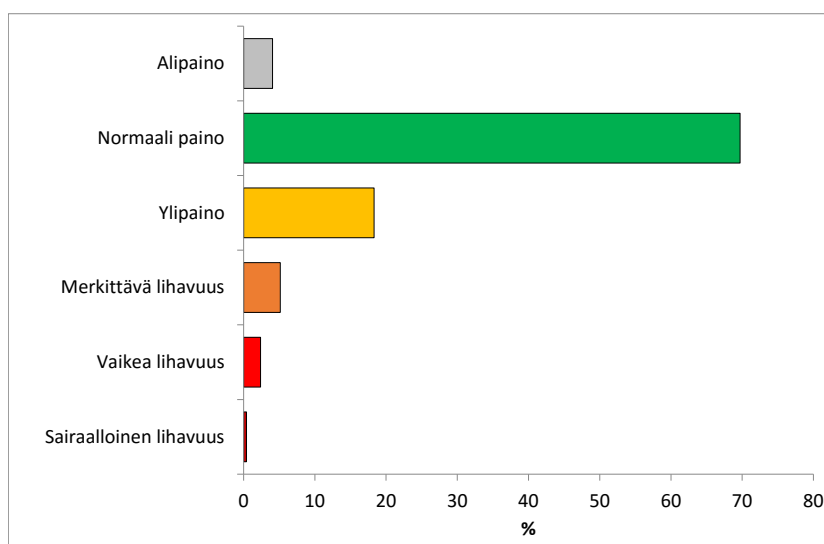
Kuvio 3.5. Äitien pituuden (cm) jakauma.

Äitien painoindeksi ennen raskautta oli keskimäärin 23.6, minimin ollessa 14.9 ja maksimin 44.6. Äitien painoissa ja painoindekseissä hajonta on siis suurta, aineistosta löytyy sekä alipainoisia että todella lihavia äitejä. Äitien ennen raskautta mitatun painoindeksin jakauma on esitetty kuviossa 3.6. Kuviosta nähdään, että painoindeksit ovat lähes normaalisti jakautuneita, mutta hieman oikealle vinoutuneita. Painoindeksien mediaani on 22.7 eli hieman pienempi kuin keskiarvo, mikä kertoo myös hieman oikealle vinoutuneesta jakaumasta.

Äidin ennen raskautta mitattu painoindeksi myös luokiteltiin kolmella tavalla kolmeksi uudeksi muuttujaksi. Näiden muuttujien teossa käytettiin Terveyskirjastossa (Mustajoki (2015)) esitettyä tapaa määrittää henkilön lihavuutta painoindeksin avulla. Muuttujissa on siis määritetty normaaliksi painoksi painoindeksi $18.5\text{--}25\text{ kg/m}^2$ ja tätä matalampi painoindeksi vastaa alipainoisuutta. Lisäksi kuusiluokkaisessa painoindeksimuuttujassa loput luokat ovat: $25.01\text{--}30\text{ kg/m}^2$ ylipaino eli lievä lihavuus, $30.01\text{--}35\text{ kg/m}^2$ merkittävä lihavuus, $35.01\text{--}40\text{ kg/m}^2$ vaikea lihavuus ja yli 40 kg/m^2 sairaalloinen lihavuus. Neliluokkaisessa painoindeksimuuttujassa kolme viimeistä luokkaa on yhdistetty yhdeksi *lihavuus* -luokaksi ja kolmeluokkaisessa



Kuvio 3.6. Äitien painoindeksin (kg/m²) jakauma.



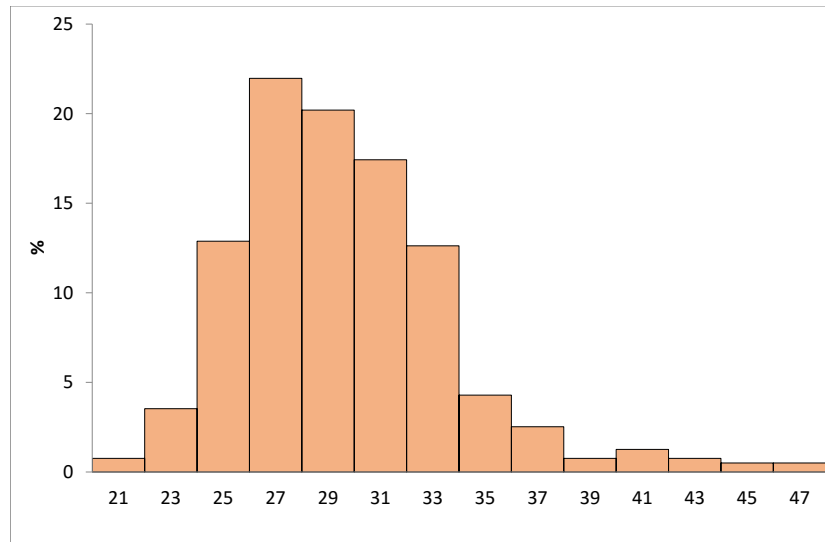
Kuvio 3.7. Kuusiluokkaisen äitien ennen raskautta mitatun painoindeksin perusteella luokitellun muuttujan jakauma.

sa muuttujassa neljä viimeistä luokkaa on yhdistetty yhdeksi *ylipaino tai lihavuus* -luokaksi. Kuviossa 3.6 on esitetty kuusiluokkaisen painoindeksimuuttujan jakauma. Siitä nähdään, että noin 70% eli suurin osa äideistä oli normaalipainoisia. 5% oli alipainoisia ja loput eli noin neljännes äideistä oli ylipainoisia tai lihavia.

Raskaudenseurannan viimeinen mittaus on joillakin äideillä tehty jo hyvin aikaisessa vaiheessa raskautta. Viimeinen raskaudenseurantamittaus oli keskimäärin tehty 269 raskauspäivän kohdalla eli 39. raskausviikolla ja aikaisimmillaan 88 raskauspäivän kohdalla eli 13. raskausviikolla. On siis selvää, että kaikki mittaukset eivät ole päteviä kuvaamaan painoindeksiä raskauden loppuvaiheilla. Raskaudenseuran-

tamittaukset myös puuttuivat kokonaan todella huomattavalta määrältä äitejä. Äidin raskaudenaikainen painoindeksi pystyttiin määrittämään vain 452:lle äidille. Näistä 396:n äidin loppuraskauden painoindeksi oli mitattu enintään kaksi viikkoa ennen synnytystä. Tätä muuttujaa tutkittaessa käytetäänkin vain näitä 396 havaintoa, joita voidaan pitää luotettavina kuvaamaan painoindeksiä raskauden loppuvaiheessa.

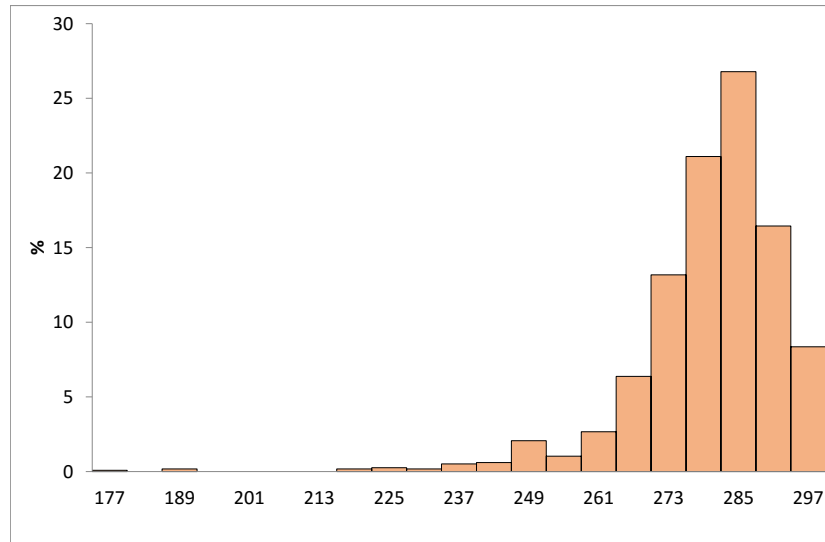
Äitien keskimääräinen paino raskaudenseurannan viimeisessä mittauksessa, joka oli tehty enintään kaksi viikkoa ennen synnytystä, oli 79 kg, pienimmän painon ollessa 55.4 kg ja suurimman 132.8 kg. Painoindeksi viimeisessä mittauksessa oli keskimäärin 28.6, minimin ollessa 19.4 ja maksimin 45.9. Hajonta painoissa ja painoindeksissä oli edelleen suurta ja mediaanit molemmissa olivat jonkin verran keskiarvoja pienempiä eli jakaumat ovat hiukan oikealle vinoutuneita. Viimeisen raskaudenseurantamittauksen painoindeksin jakauma on esitetty kuviossa 3.8. Myös jakaumakuviointi sisältää vain ne painoindeksimittaukset, jotka oli tehty enintään kaksi viikkoa ennen synnytystä.



Kuvio 3.8. Äitien viimeisen raskaudentaseurantamittauksen painoindeksin (kg/m^2) jakauma. Jakaumassa mukana vain ne äidit, joiden painoindeksimittaus oli tehty enintään kaksi viikkoa ennen synnytystä.

Äitien keskimääräinen raskauden kesto oli 278 päivää eli noin 39 viikkoa ja 5 päivää. Tämä on lähellä normaaliksi määriteltyä raskauden kestoja, joka on 280 päivää eli 40 viikkoa (Tiitinen 2015). Raskauden keston mediaani oli 280 eli hieman suurempi kuin keskiarvo. Jakauma olikin jossain määrin vasemmalle vino. Raskauden keston jakautuminen on esitetty kuviossa 3.9. Kuvioista nähdään, että aineistosta löytyy joitakin selvästi aliaikaisia raskauksia, joista erityisesti pari on selvästi poikkeavia havaintoja. Tieto raskauden kestoista löytyi 1161:ltä äidiltä eli lähes kaikilta.

Synnyttäneiden äitien keskimääräinen ikä oli 29.9 vuotta, kun tarkasteltiin äitien ikää vuoden tarkkuudella. Äideille laskettiin myös tarkemmat iät päivän tarkkuudella lapsen syntymäpäivän ja äidin syntymäpäivän erotuksena. Tällöin iän keskiarvoksi tuli 30.4 vuotta (11110 päivää). Vanhin synnyttänyt äiti oli 45-vuotias ja nuorin 15-



Kuvio 3.9. Äitien raskauden keston (pv) vasemmalle vinoutunut jakauma.

vuotias. Äitien ikä noudatti täysin normaalijakaumaa. Kaikilta äideiltä löytyi tieto iästä.

Suurin osa äideistä (90.6 %) ei tupakoinut raskauden aikana. Raskauden aikana tupakoineista 1.4 % lopetti tupakoinnin raskauden ensimmäisen kolmanneksen aikana ja loput jatkoivat vielä tämän jälkeen. 1134:lle äidille löytyi tieto raskauden aikaisesta tupakoinnista. Suurin osa äideistä ei myöskään sairastanut diabetesta (81.55 %). Tyypin I diabetesta sairasti 1.45 % ja raskauden aikaiseen diabetekseen sairastui 17 % äideistä. 1166:lle äidille oli saatavilla tieto diabeteksestä.

Tunnusluvut liittyen äitien aiempiin raskauksiin, raskauden keskeytyksiin, keskenmenoihin, kohdun ulkopuolisiin raskauksiin, aiempiin synnytyksiin sekä kuolleena syntyneisiin lapsiin on esitetty taulukossa 3.1. 47 % aineiston äideistä oli ensisynnyttäjiä ja yhteensä 94 prosentilla synnytys oli enintään kolmas. 36 % ei ollut ollut aiemmin raskaana. 78 % ei ollut saanut ikinä keskenmenoa, 21 % oli saanut yhden tai kaksi keskenmenoa. 86 % ei ollut tehnyt aiemmin raskaudenkeskeytystä, 13 % oli tehnyt yhden tai kaksi keskeytystä. 99 prosentilla ei ollut ollut kohdun ulkopuolisia raskauksia. Vain kolme äitiä oli aiemmin synnyttänyt kuolleen lapsen.

Taulukko 3.1. Äitien aiempiin raskauksiin liittyviä tunnuslukuja.

variable	N	Keskiarvo	Keskihajonta	Min	Max
aiemmat raskaudet	1166	1.3	1.4	0	10
keskenmenot	1166	0.3	0.6	0	5
keskeytykset	1166	0.2	0.5	0	3
kohdun ulkopuoliset raskaudet	1166	0.0	0.1	0	1
aiemmat synnytykset	1166	0.8	1.0	0	8
kuolleena syntyneet	1166	0.0	0.1	0	1

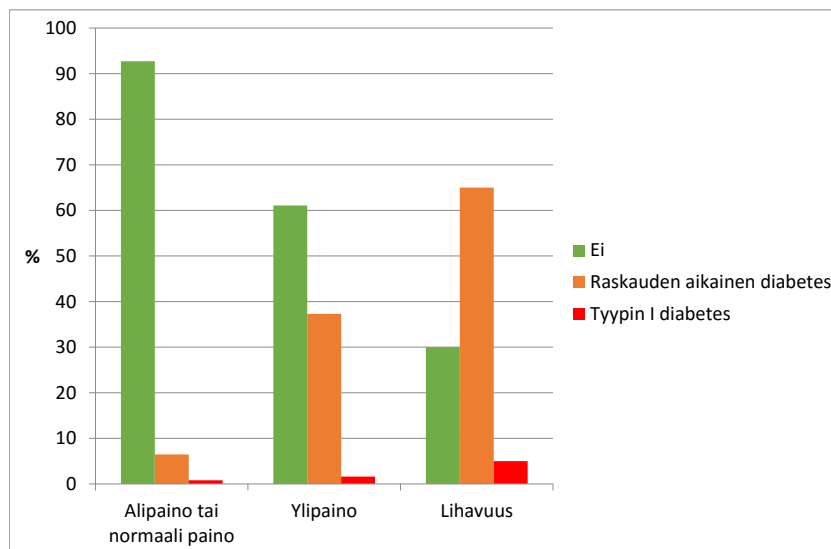
Äideistä 84.3 % imetti lastaan synnytyssairaalasta lähdettäessä kerätyn tiedon

mukaan ja myöhemmin synnytyksen jälkitarkastuksessa kerätyn tiedon mukaan imettävien äitien osuus oli 71.3 %. Lisäksi lastaan osittain imettävien vastaavat osuudet olivat 12.9% ja 16.3 %. Loput äidit eivät imettäneet lastaan ollenkaan kyseisinä ajankohtina. Näissä muuttujissa oli kuitenkin huomattava määrä puuttuvaa tietoa, tieto imetyksestä löytyi sairaalasta lähettäessä 420:ltä äidiltä ja synnytyksen jälkitarkastuksessa 435:ltä äidiltä eli selvästi alle puolilta kaikista äideistä.

3.2.3 Selittävien muuttujien väliset yhteydet

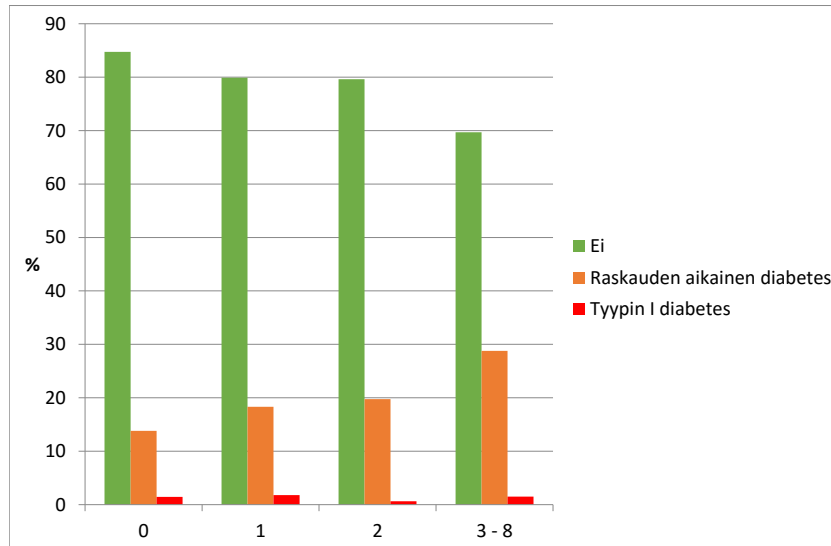
Tässä alaluvussa on esitetty tässä tutkielmassa käytettävien selittävien muuttujien välisiä yhteyksiä. Tällaisista mahdollisista yhteyksistä on hyvä olla tietoinen malleja rakentaessa, sillä keskenään korreloivat muuttujat voivat kumota toistensa vaikutuksen ollessaan samanaikaisesti mallissa selittäjinä.

Äidin painoindeksi ja diabetes olivat odotetustikin yhteydessä toisiinsa. Tätä havainnollistaa kuvio 3.10. Kuviossa on yhdistettynä alipainoiset ja normaalipainoiset äidit, sillä näissä ryhmissä diabeteksen jakaumat olivat lähes identtiset keskenään. Kuviosta nähdään, että diabeetikoiden osuudet ovat selvästi erilaiset riippuen painoindeksistä. Ylipainoisissa äideissä raskauden aikaisen diabeteksen osuus on moninkertaisesti suurempi kuin ali- tai normaalipainoissa, ja lihavissa äideissä osuus suurenee vielä entisestään. Lihavista äideistä jopa noin 65 % sairastui raskauden aikana diabetekseen. Myös tyypin I diabeteksen osuus suurenee painoindeksin suuren tuessa. Tyypin I diabeetikkoja tässä tarkastelussa oli kuitenkin vain 13, joten tämän pohjalta ei voida tehdä mitään päätelmiä.



Kuvio 3.10. Diabeteksen jakautuminen äidin ennen raskautta mitatun painoindeksin mukaan.

Diabeetikoiden jakaumat erosivat myös äidin aiempien synnytysten lukumäärän mukaan. Kuviosta 3.11 voidaan nähdä, että diabeetikoiden määrä sitä suurempi mitä useampi synnytys äidillä on takana. Ensisyntyttäjistä noin 85 % ei sairastanut diabetesta, kun taas 3-8 kertaa aiemmin synnyttäneistä vastaava osuus on noin 70 %.



Kuvio 3.11. Diabeteksen jakautuminen äidin aiempien synnytysten lukumäärän mukaan.

Taulukossa 3.2 on esitetty tutkielman tärkeimpien kvantitatiivisten selittävien muuttujien väliset korrelaatiokerroimet. Korrelaatiokerrointen merkitsevyyttä testattiin kaksisuuntaisella korrelaatiotestillä. Testin hypoteesit olivat

$$H_0 : r = 0 \text{ ja } H_1 : r \neq 0.$$

Testausta varten laskettiin t-testisuure

$$t = r \sqrt{\frac{n-2}{1-r^2}},$$

joka noudattaa Studentin t-jakaumaa vapausastein $n-2$ nollahypoteesin toteutuksessa. Kaavassa r on korrelaatiokerroin ja n on havaintojen lukumäärä. Taulukossa 3.2 korrelaatiokerroin on keltaisella pohjalla, jos se oli erittäin merkitsevä eli $p < 0.001$, tummennettuna, jos se oli merkitsevä eli $p < 0.01$ ja alleviivattuna, jos se oli melkein merkitsevä eli $p < 0.05$.

Jotkut korrelaatiot olivatkin odotettuja, kuten lapsen syntymämittausten väliset vahvat korrelaatiot. Äidin ennen raskautta mitatun painoindeksin ja lapsen syntymämittausten välillä mikään korrelaatio ei ollut tilastollisesti suuresti merkitsevää.

Taulukko 3.2. Korrelaatiomatriisi, jossa esitetty aineiston kvantitatiivisten selittävien muuttujien väliset korrelaatiokertoimet ja niiden merkitsevyys; keltainen: $p < 0.001$, tummennus: $p < 0.01$ ja alleviivaus: $p < 0.05$.

	syntymäpaino	syntymäpituus	syntymäBMI	päännympäryys	äidin ikä	aiemmat synnytykset	äidin paino	äidin pituus	äidin BMI	raskauden kesto (pv)	äidin paino (raskaus)	äidin BMI (raskaus)	painon muutos	BMI muutos
syntymäpaino	1.00	0.80	0.84	0.71	-0.01	0.17	0.10	0.15	0.05	0.63	0.20	0.11	0.19	0.14
syntymäpituus	0.80	1.00	0.37	0.61	0.01	0.10	0.09	0.15	0.04	0.64	0.24	0.17	0.14	0.09
syntymäBMI	0.84	0.37	1.00	0.55	-0.04	0.17	0.04	0.11	0.00	0.44	0.07	0.01	0.13	0.11
päännympäryys	0.71	0.61	0.55	1.00	0.09	0.19	0.13	0.16	0.08	0.43	0.17	0.08	0.05	0.00
äidin ikä	-0.01	0.01	-0.04	0.09	1.00	0.38	0.12	0.05	0.12	-0.09	-0.01	-0.02	-0.22	-0.23
aiemmat synnytykset	0.17	0.10	0.17	0.19	0.38	1.00	0.07	-0.04	0.10	0.03	0.04	0.05	-0.16	-0.16
äidin paino	0.10	0.09	0.04	0.13	0.12	0.07	1.00	0.36	0.92	-0.07	0.91	0.82	-0.20	-0.27
äidin pituus	0.15	0.15	0.11	0.16	0.05	-0.04	0.36	1.00	-0.02	0.03	0.41	-0.04	0.10	-0.09
äidin BMI	0.05	0.04	0.00	0.08	0.12	0.10	0.92	-0.02	1.00	-0.09	0.82	0.90	-0.26	-0.26
raskauden kesto (pv)	0.63	0.64	0.44	0.43	-0.09	0.03	-0.07	0.03	-0.09	1.00	-0.01	-0.03	0.17	0.16
äidin paino (raskaus)	0.20	0.24	0.07	0.17	-0.01	0.04	0.91	0.41	0.82	-0.01	1.00	0.89	0.22	0.14
äidin BMI (raskaus)	0.11	0.17	0.01	0.08	-0.02	0.05	0.82	-0.04	0.90	-0.03	0.89	1.00	0.18	0.19
painon muutos	0.19	0.14	0.13	0.05	-0.22	-0.16	-0.20	0.10	-0.26	0.17	0.22	0.18	1.00	0.97
BMI muutos	0.14	0.09	0.11	0.00	-0.23	-0.16	-0.27	-0.09	-0.26	0.16	0.14	0.19	0.97	1.00

Raskaudenseurannan viimeisen mittauksen paino ja painoindeksi puolestaan korreloivat merkitsevästi joidenkin lapsen syntymämittausten kanssa. Myös äidin painon ja painoindeksin raskaudenaikaisen muutoksen ja lapsen syntymämittausten välillä oli joitakin merkitseviä korrelaatioita. Äidin pituus korreloi positiivisesti erittäin merkitsevästi lapsen kaikkien syntymämittausten kanssa. Äidin ikä korreloi positiivisesti äidin painon ja painoindeksin kanssa sekä negatiivisesti raskauden aikaisen painon- ja painoindeksin muutoksen kanssa. Vanhemmillä äideillä painoa on siis enemmän, mutta sitä kerääntyy raskauden aikana lisää vähemmän kuin nuoremmilla äideillä. Äitien aiempien synnytysten lukumäärä korreloi positiivisesti lapsen kaikkien syntymämittausten kanssa. Tämän perusteella lapset olisivat isompia, jos äiti on jo synnyttänyt aiemmin. Äidin ikä ja raskauden kesto korreloivat negatiivisesti, joten iäkkäämmillä äideillä raskauden kesto oli hieman lyhyempi kuin nuoremmilla. Raskauden kesto luonnollisesti korreloi vahvasti positiivisesti lapsen syntymämittausten kanssa.

4 Tutkimuksessa käytetyt menetelmät

Seuraavissa alaluvuissa on esitetty tutkielmassa käytettyjä tilastollisia menetelmiä. Alaluvussa 4.1 on esitetty tasoittavat splinit ja sen alaosiossa 4.1.1 luonnolliset kuutiotasoitettut splinit, joita on käytetty painoindeksikäyrien mallinnuksessa 2 ja 10 ikävuoden välillä. Splinikäyriä on estimoitu painoindeksin käännepisteet. Alaosiossa 4.1.2 on esitetty kuutiotasoitettujen splinien tasoitusparametrin valintaan käytetty yleistetty ristiinvalidointi. Alaluvussa 4.2 on esitetty lineaarinen regressiomalli, jonka avulla on tutkittu painoindeksin käännepisteeseen yhteydessä olevia tekijöitä. Alaosiossa 4.2.1 on esitetty lineaariseen regressioanalyysiin liittyvä hypoteesien testaus ja alaosiossa 4.2.2 polynomiregressiomalli, jota on käytetty painoindeksien mallinnuksessa 0 ja 3 ikävuoden välillä. Neljännen asteen polynomiregressiomallien avulla on estimoitu painoindeksin varhaislapsuuden piikit. Alaluvussa 4.3 on esitetty lineaarinen sekoitusregressiomalli. Sen avulla on tutkittu painoindeksin käännepisteeseen yhteydessä olevia tekijöitä. Alaosiossa 4.3.1 on esitetty Akaiken informaatiokriteeri, jota on käytetty tutkielman mallien valintaan ja vertailuun, erityisesti sekoitusregressiomallin valintaan.

4.1 Tasoittava splini

Tämä alaluku ja sen ensimmäinen alaosio perustuvat kirjoihin Nonparametric Curve Estimation: Methods, Theory & Applications (Efromovich (1999)) ja Generalized Additive Models: An introduction with R (Wood (2006)).

Splinit ovat paloittaisia käyriä, jotka yhdistyvät solmukohdissa yhdeksi jatkuvaksi käyräksi. Splineissa käytettäviä tasoitusfunktioita on erilaisia, mutta ne voidaan yleisesti esittää mallina, joka sisältää yhden kovariaatin yhden tasoitusfunktion. Tällainen malli on muotoa

$$(4.1) \quad y_i = f(x_i) + \epsilon_i, \quad i = 1, 2, \dots, n,$$

missä y_i on riippuva vastemuuttuja, f on tasoitusfunktio, x_i on kovariaatti ja ϵ_i :t ovat riippumattomia ja identtisesti $N(0, \sigma^2)$ jakautuneita satunnaismuuttujia. Oletetaan, että kovariaatit x_i ovat välillä $[a, b]$.

Tasoittava splini on lineaarinen estimaattori, joten sovitettavat arvot \hat{y}_i voidaan kirjoittaa myös vektorimuodossa

$$(4.2) \quad \hat{\mathbf{y}} = A(\lambda)\mathbf{y},$$

missä $\hat{\mathbf{y}}$ on sovitettujen arvojen vektori, \mathbf{y} on havaittujen arvojen vektori ja $A(\lambda)$ on tasoittajamatriisi

$$(4.3) \quad A(\lambda) = X(X^T X)^{-1} X^T,$$

joka muuntaa havaittujen arvojen vektorin sovitettujen arvojen vektoriksi. Tasoittajamatriisin elementit $A(\lambda) = A_{ij}(\lambda)$ toteuttavat seuraavat ehdot

$$0 \leq A_{ii}(\lambda) \leq 1,$$

$$-1 \leq A_{ij}(\lambda) \leq 1, \text{ kaikille } i \neq j.$$

Tasoittajamatriisin diagonaalelementit $A_{ii}(\lambda)$ kuvaavat jokaisen havaitun arvon vaikutusta saman havainnon sovitettuun arvoon. Kaikki diagonaalelementit voivat saada arvokseen 1 vain jos matriisin kaikki muut elementit ovat 0.

Tasoittavan splinen estimaattori saadaan laskettua minimoimalla neliösumma

$$(4.4) \quad \sum_{i=1}^n \{y_i - f(x_i)\}^2 + \lambda \int_a^b \{f^{(m)}(x)\}^2 dx,$$

missä λ on annettu tasoitusparametri, jonka avulla määritetään splinikäyrän taiseisuus ja aineistoon sopivuus. Jos tasoitusparametri on lähellä nollaa, johtaa se tilanteeseen, missä splinikäyrä kulkee tarkalleen kaikkien mittapisteiden kautta. Jos taas tasoitusparametri lähestyy ääretöntä, käyrä lähestyy regressiosuoraa. Toisin sanoen, jos tasoitusparametri on liian suuri, mallista voi tulla ylitasoitettu ja jos se taas on liian pieni, mallista voi tulla alitasoitettu. Kaavan (4.4) ratkaisu on siis tasoittava splini, jossa x_i ovat solmukohtia. Kaavassa ensimmäinen termi määrittää aineistolle sovitettujen splinien tarkkuutta ja toinen termi määrittää splinin tasoituksen.

4.1.1 Luonnollinen kuutiotasoitettu splini

Tässä tutkielmassa painoindeksikäyriä mallinnetaan kuutiotasoitetuilla splineilla eli aineistoon sovitetaan tasoitettu käyrä, joka kulkee lapsen mittapisteiden läpi. Kuutiotasoituksella tarkoitetaan sitä, että splinikäyrät ovat paloittain määriteltyjä kolmannen asteen polynomifunktioita solmukohtien väleillä. Tämän menetelmän avulla käyrä saadaan sovitettua hyvin aineistoon, vaikka havaintojen arvot vaihtelisivatkin paljon.

Solmukohtien väleille määritellyt funktiot ovat muotoa

$$(4.5) \quad f(x) = \beta_1 + \beta_2 x + \beta_3 x^2 + \beta_4 x^3.$$

Tällöin siis kaava (4.1) voidaan esittää muodossa

$$(4.6) \quad y_i = \beta_1 + \beta_2 x_i + \beta_3 x_i^2 + \beta_4 x_i^3 + \epsilon_i, \quad i = 1, 2, \dots, n.$$

Kuutiotasoitettun splinen estimaattori saadaan laskettua asettamalla $m=2$ pienimmän neliösumman kaavaan (4.4).

4.1.2 Tasoitusparametrin λ valinta

Tasoitusparametrin valinta on tärkeä osa mallin estimointia. Tasoitusparametrin valinnan voi suorittaa joko silmämääräisesti valitsemalla sopiva arvo tai sitten tähän tarkoitukseen soveltuvilla automaattisilla menetelmillä eli esimerkiksi ristiinvalidoinnilla (CV) tai yleistetyllä ristiinvalidoinnilla (GCV). Tässä tutkielmassa on käytetty lasten painoindexikäyrien mallinnukseen näistä jälkimmäistä eli yleistettyä ristiinvalidointia. Tässä alaosiossa on kerrottu tarkemmin ristiinvalidoinnista sekä yleistetystä ristiinvalidoinnista. Alaosio perustuu kirjoihin Generalized Additive Models: An introduction with R (Wood (2006)) ja Smoothing Methods in Statistics (Simonoff (2012)).

Määritellään ensin tavallinen ristiinvalidointi, jonka pohjalta määritellään yleistetty ristiinvalidointi. Molemmat ovat siis tasoitusparametrin valintaan soveltuvia menetelmiä. Tavallisessa ristiinvalidoinnissa tasoitusparametri valitaan minimoimalla ristiinvalidoinnin pistemääräfunktiota

$$(4.7) \quad CV(\lambda) = \frac{1}{n} \sum_{i=1}^n \{y_i - \hat{f}_\lambda^{(-i)}(x_i)\}^2, \quad i = 1, \dots, n,$$

missä $\hat{f}_\lambda^{(-i)}(x_i)$ on spliniestimaatti pisteessä x_i , kun malli on sovitettu koko muulle aineistolle lukuunottamatta havaintopistettä y_i . Ristiinvalidoinnin avulla valittu λ :n arvo on siis se, jolla kaava (4.7) saadaan minimoitua. Tämän funktion laskenta on kuitenkin työlästä, sillä kaikki lokaalit sovitteet $\hat{f}_\lambda^{(-i)}(x_i)$ on laskettava erikseen. Voidaankin osoittaa, että kaava (4.7) voidaan muuttaa yksinkertaisempaan muotoon kirjoittamalla se sovitettujen arvojen funktiona

$$(4.8) \quad CV(\lambda) = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{y_i - \hat{f}_\lambda(x_i)}{1 - A_{ii}(\lambda)} \right\}^2 = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{y_i - \hat{y}_i}{1 - A_{ii}(\lambda)} \right\}^2,$$

missä $A_{ii}(\lambda)$ on tasoittajamatriisin i :nnes diagonaalelementti ja \hat{y}_i on koko aineistolle sovitettu spliniestimaatti.

Tästä päästään yleistetyn ristiinvalidoinnin kaavaan, jossa kaavan (4.8) kaikki tasoittajamatriisin diagonaalelementit $A_{ii}(\lambda)$ korvataan niiden keskiarvolla eli

$$(4.9) \quad \frac{1}{n} \sum_{i=1}^n A_{ii}(\lambda) = n^{-1} \text{tr}[A(\lambda)].$$

Tasoittajamatriisin diagonaalelementtien summa $\text{tr}[A(\lambda)]$ on myös sovitetun splinimallin vapausasteiden lukumäärä (df). Vapausasteiden lukumäärän avulla voidaan laskea, kuinka monta estimoitua parametria splinimallissa on: $\text{df} = n - p$, missä n on havaintojen lukumäärä ja p on estimoitujen parametrien lukumäärä.

Nyt tasoitusparametri voidaan valita minimoimalla yleistetyn ristiinvalidoinnin pistemääräfunktiota

$$(4.10) \quad GCV(\lambda) = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{y_i - \hat{y}_i}{1 - n^{-1} \text{tr}[A(\lambda)]} \right\}^2.$$

Ristiinvalidoinnin ja yleistetyn ristiinvalidoinnin eroa voidaan havainnollistaa kirjoittamalla kaava (4.10) muotoon

$$(4.11) \quad GCV(\lambda) = \frac{1}{n} \sum_{i=1}^n \left(\left\{ \frac{1 - A_{ii}(\lambda)}{1 - n^{-1} \text{tr}[A(\lambda)]} \right\}^2 [y_i - \hat{f}_\lambda^{(i)}(x_i)]^2 \right).$$

Kun verrataan kaavaa (4.11) kaavaan (4.7), on helppo huomata, että jos tasoittajamatriisin diagonaalelementtien $A_{ii}(\lambda)$ arvot ovat kaikki samoja, nämä kaksi menetelmää antavat saman tasoitusparametrin arvon. Muussa tapauksessa yleistetty ristiinvalidointi vähentää vahvasti vaikuttavien havaintopisteiden vaikutusta tasoitusparametrin valintaan. Toisin sanoen, kun havainnot ovat sijoittuneet hyvin epätasaisesti, yleensä yleistetty ristiinvalidointi on paremmin toimiva menetelmä tasoitusparametrin valintaan.

4.2 Lineaarinen regressiomalli

Tämä alaluku ja sen alaosiot perustuvat kirjoihin Introduction to Linear Regression Analysis (Montgomery, Peck & Vining (2006)) ja Applied Linear Regression (Weisberg (2013)). Lineaarisen regressiomallin avulla voidaan tutkia muuttujien välistä lineaarista riippuvuutta. Mallissa selitettävää muuttujaa merkitään y :llä ja selittäviä muuttujia x_i :llä, missä selittävien muuttujien lukumäärä $i = 1, 2, \dots, k$. Selittäviä muuttujia voi siis olla yksi tai useampia ja ne voivat olla kvantitatiivisia tai kvalitaatiivisia dummy-muuttujia. Lineaarinen regressiomalli on muotoa

$$(4.12) \quad y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \epsilon,$$

missä β_0 on vakiotermi, β_k ovat selittävien muuttujien regressiokertoimia ja ϵ on mallin virhetermi. Kun $k=1$, mallissa on vain yksi selittävä muuttuja ja silloin kyse on yksinkertaisesta lineaarisesta regressiosta. Lineaarisessa regressiomallissa oletuksena on, että virhetermit noudattavat jakaumaa $N(0, \sigma^2)$ eli virhetermien odotusarvo on nolla ja varianssi σ^2 . Virhetermit eivät korreloi keskenään. Lisäksi myöskään selittävät muuttujat eivät korreloi keskenään eli ne ovat lineaarisesti riippumattomia toisistaan.

Lineaarinen regressiomalli voidaan esittää myös matriisimuodossa:

$$(4.13) \quad \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

missä

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix}, \mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{12} \dots x_{1k} \\ 1 & x_{21} & x_{22} \dots x_{2k} \\ \dots & \dots & \dots \\ 1 & x_{n1} & x_{n2} \dots x_{nk} \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_k \end{pmatrix}, \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \dots \\ \epsilon_n \end{pmatrix},$$

missä n on havaintojen lukumäärä ja k mallin selittävien muuttujien lukumäärä. Mallissa (4.13) \mathbf{y} on selitettävän muuttujan $n \times 1$ havaintovektori, \mathbf{X} on selitettävien muuttujien arvojen $n \times p$ matriisi, $\boldsymbol{\beta}$ on regressiokertoimien $p \times 1$ vektori ja $\boldsymbol{\epsilon}$ on virhetermien $n \times 1$ vektori.

Mallin parametrien estimointi tapahtuu pienimmän neliösumman menetelmällä eli etsitään mallin parametrien vektori $\hat{\boldsymbol{\beta}}$, joka minimoi residuaalien neliösumman

$$(4.14) \quad S(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).$$

Kaava (4.14) voidaan esittää myös muodossa

$$(4.15) \quad \begin{aligned} S(\boldsymbol{\beta}) &= \mathbf{y}'\mathbf{y} - \boldsymbol{\beta}'\mathbf{X}'\mathbf{y} - \mathbf{y}'\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \\ &= \mathbf{y}'\mathbf{y} - 2\boldsymbol{\beta}'\mathbf{X}'\mathbf{y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}. \end{aligned}$$

Pienimmän neliösumman estimaattorin täytyy toteuttaa yhtälö

$$(4.16) \quad \left. \frac{\partial S}{\partial \boldsymbol{\beta}} \right|_{\hat{\boldsymbol{\beta}}} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{0},$$

mistä saamme $\boldsymbol{\beta}$:n pienimmän neliösumman estimaattorin

$$(4.17) \quad \hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}.$$

Kaavan (4.17) käänteismatriisi $(\mathbf{X}'\mathbf{X})^{-1}$ on olemassa aina, kun selittävät muuttujat ovat lineaarisesti riippumattomia. Tämä on yksi lineaarisen mallin oletuksista. Sovitettu malli on nyt muotoa

$$(4.18) \quad \hat{\mathbf{y}} = (\mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{H}\mathbf{y},$$

missä \mathbf{H} on tasoittajamatriisi, joka on määritelty jo kaavojen (4.2) ja (4.19) yhteydessä.

4.2.1 Hypoteesien testaus

Lineaarisen regressiomallin on tärkeää olla tilanteeseen sopiva. Mallin sopivuutta voidaan tutkia esimerkiksi testaamalla regression merkitsevyyttä eli sitä, onko selittävän muuttujan y ja edes yhden selittävän muuttujan x_1, x_2, \dots, x_k välillä olemassa lineaarinen yhteys. Tällöin hypoteesit ovat:

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_1 : \beta_j \neq 0 \text{ ainakin yhdellä } j:n \text{ arvolla } 1, 2, \dots, k.$$

Nollahypoteesin hylkäys tarkoittaa käytännössä sitä, että ainakin yksi selittävästä muuttujasta x_1, x_2, \dots, x_k vaikuttaa merkitsevästi malliin. Testisuuren laskemista varten kokonaisneliösumma jaetaan kahteen osaan, regressioneliösummaan ja jäännöseliösummaan:

$$SS_T = SS_R + SS_{Res}.$$

Tätä kutsutaan varianssianalyysihajotelmaksi. Siinä kokonaisneliösumma SS_T mittaa selittävän muuttujan y kokonaisvaihtelua ja se voidaan siis jakaa regressiösummaan SS_R , joka mittaa estimoitujen \hat{y} arvojen vaihtelua, ja jäännöseliösummaan SS_{Res} , joka mittaa mallin residuaalien vaihtelua. SS_R kuvaa sitä osuutta selittävän muuttujan y vaihtelusta, jonka regressiomalli on selittänyt, ja SS_{Res} sitä osuutta, jota se ei ole pystynyt selittämään. Neliösummien kaavat ovat muotoa

$$SS_R = \|\hat{\mathbf{y}} - \bar{\mathbf{y}}\|^2$$

$$SS_{Res} = \|\mathbf{y} - \hat{\mathbf{y}}\|^2$$

Jos nollahypoteesi jää voimaan, tällöin

$$F = \frac{SS_R/k}{SS_{Res}/(n-k-1)} = \frac{MS_R}{MS_{Res}}$$

noudattaa $F_{k, n-k-1}$ jakaumaa. Tässä k on selittävien muuttujien lukumäärä ja n havaintojen lukumäärä. Nollahypoteesia voidaan testata F-testillä ja se voidaan hylätä, jos testisuure

$$F > F_{\alpha; k, n-k-1}, \text{ missä } \alpha \text{ on riskitaso.}$$

Mallin sopivuutta voidaan tutkia myös määrittämällä mallin selitysaste (R^2) ja korjattu selitysaste (R_{adj}^2). Mallin selitysaste kuvaa sitä, kuinka suuren osan mallin selittävät muuttujat selittävät y :n kokonaisvaihtelusta. Usein vertailtaessa malleja toisiinsa käytetään korjattua selitystasetta. Selitystasteen ja korjatun selitystasteen kaavat ovat muotoa

$$R^2 = 1 - \frac{SS_{Res}}{SS_T}$$

$$R_{adj}^2 = 1 - \frac{SS_{Res}/(n-p)}{SS_T/(n-1)},$$

missä n on havaintojen lukumäärä ja p estimoitujen parametrien lukumäärä. Korjatussa selityksasteessa kaavan alapuoli pysyy vakiona riippumatta siitä, kuinka monta selittävää muuttujaa mallissa on. Kaavoista saatava luku on väliltä $[0, 1]$ ja se voidaan ilmoittaa myös prosenttilukuna.

Jos saadaan selville, että ainakin joku selittävistä muuttujista on mallissa merkitsevä eli malli on sopiva, voidaan testata, mikä tai mitkä muuttujat ovat kyseessä. Kun testataan yksittäisten selittävien muuttujien merkitsevyyttä mallissa, hypoteesit ovat:

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0.$$

Jos nollahypoteesi jää voimaan, tarkoittaa se sitä, että selitettävä muuttuja y ei riipu lineaarisesti kyseisestä yksittäisestä selittävästä muuttujasta. Tämä selittävä muuttuja ei ole siis mallissa merkitsevä ja se voidaan jättää mallista pois. Nollahypoteesin ollessa voimassa

$$t = \frac{\hat{\beta}_j}{s(\hat{\beta}_j)}$$

noudattaa $t(n-k-1)$ jakaumaa. Tässä $s(\hat{\beta}_j)$ on estimoidun parametrin hajonta, n on havaintojen lukumäärä ja k on mallin selittävien muuttujien lukumäärä. Nollahypoteesi voidaan hylätä riskitasolla α , jos $|t| > t_{\alpha/2, n-k-1}$. On tärkeää muistaa kuitenkin, että estimoidut parametrit $\hat{\beta}_j$ riippuvat myös muista mallin selittävistä muuttujista $x_i (i \neq j)$. Voidaan siis ajatella, että testataan selittävän muuttujan x_j osuuden merkitsevyyttä kaikista mallin selittäjistä.

4.2.2 Polynomiregressiomalli

Polynomiregressiomalli on erityistapaus lineaarisesta regressiomallista (4.12), mutta selittävän muuttujan x ja selitettävän muuttujan y välistä yhteyttä mallinnetaan n :nnen asteen polynomin avulla. Polynomiregressiomallin avulla voidaan tutkia $x:n$ ja $y:n$ välistä epälineaarista riippuvuutta, vaikka malli voidaanakin tulkita lineaariseksi.

Yhden selittävän muuttujan polynomiregressiomallin kaava on muotoa

$$(4.19) \quad y = \beta_0 + \beta_1 x + \beta_2 x^2 + \dots + \beta_a x^a + \epsilon,$$

missä y on selitettävä muuttuja, x on selittävä muuttuja ja a on polynomin aste. Tällöin $y:n$ odotusarvo

$$E(y) = \beta_0 + \beta_1 x + \beta_2 x^2 + \dots + \beta_a x^a.$$

Yhden selittäjän polynomiregressiomalli voidaan kirjoittaa matriisimuotoon kaavan (4.13) avulla niin, että

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix}, \mathbf{X} = \begin{pmatrix} 1 & x & x^2 & \dots & x^a \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_a \end{pmatrix}, \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \dots \\ \epsilon_n \end{pmatrix},$$

missä n on havaintojen lukumäärä ja a polynomin aste. Mallin parametrien estimointi tapahtuu pienimmän neliösumman menetelmällä kaavoissa (4.15), (4.16) ja (4.17) esityllä tavalla.

4.3 Lineaarinen sekoitusregressiomalli

Tämä alaluku perustuu kirjaan Finite Mixture Models (McLachlan & Peel (2004)) ja artikkeliin FlexMix: A General Framework for Finite Mixture Models and Latent Class Regression in R (Leisch (2004)). Lineaarisen sekoitusregressiomallin avulla voidaan tutkia muuttujien välisiä lineaarisia yhteyksiä aineiston aliryhmissä eli komponenteissa. Sekoitussmallien ideana on jakaa aineisto komponentteihin eli etsiä aineistosta niin sanottuja piileviä ryhmiä ja sovittaa jokaiselle komponentille oma mallinsa. Näin voidaan tutkia, onko aineistossa aliryhmiä, joissa selittävien muuttujien vaikutus selitettävään muuttujaan on keskenään erilaista. Sekoitusmalleja voidaan muodostaa muistakin malleista kuin lineaarisesta regressiomallista.

Sekoitusmallin jakauman tiheysfunktio on muotoa

$$(4.20) \quad f(y) = \sum_{g=1}^G \pi_g f_g(y),$$

missä G on komponenttien lukumäärä ja π_g ovat epänegatiivisia sekoituksen painoarvoja, jotka ovat osuuksia väliltä

$$0 \leq \pi_g \leq 1.$$

Nämä osuudet summautuvat ykköseen eli

$$\sum_{g=1}^G \pi_g = 1.$$

Voidaan sanoa, että jakauma f on sekoitus jakaumista f_1, f_2, \dots, f_G ja sisältää G komponenttia.

Lineaarisen sekoitusregressiomallin kaava

$$(4.21) \quad y_{(g)} = \beta_{0(g)} + \beta_{1(g)}x_1 + \beta_{2(g)}x_2 + \dots + \beta_{k(g)}x_k + \epsilon_{(g)}, \quad g = 1, 2, \dots, G$$

näyttää lähes samalta kuin tavallisen lineaarisen regressiomallin kaava (4.12) lu-
kuunottamatta alaindeksien g -kirjaimia. Alaindeksi g kertoo, minkä piilevän luokan

sovitettu lineaarinen regressiomalli on kyseessä, G on piilevien luokkien kokonaismäärä ja k on mallin selittävien muuttujien lukumäärä.

Posterioritodennäköisyys sille, että tietty havainto (x, y) kuuluu luokkaan j on

$$(4.22) \quad P(j|x, y, \psi) = \frac{\pi_j f(y|x, \theta_j)}{\sum_g \pi_g f(y|x, \theta_g)},$$

missä $g=1, 2, \dots, G$ on mallin komponentti. Uskottavuusfunktion logaritmi muuttujajoukolle $\{(x_1, y_1), \dots, (x_n, y_n)\}$ on

$$(4.23) \quad \log L = \sum_{i=1}^n \log f(y_i|x_i, \psi) = \sum_{i=1}^n \log \left(\sum_{g=1}^G \pi_g f(y_i|x_i, \theta_g) \right),$$

missä n on havaintojen lukumäärä ja G komponenttien lukumäärä. Uskottavuusfunktiota ei yleensä pystytä suoraan maksimoimaan, joten parametrivektorin ψ estimoinnissa käytetään apuna EM (expectation maximization)-algoritmia. Algoritmi sisältää kaksi vaihetta, niin sanotut E- ja M-askeleet.

E-askel: Estimoidaan kaavan (4.22) avulla posterioritodennäköisyydet jokaiselle havainnolle

$$\hat{p}_{ig} = P(g|x_i, y_i, \hat{\psi})$$

Posterioritodennäköisyyksien avulla voidaan laskea prioritodennäköisyydet

$$\hat{\pi}_g = \frac{1}{n} \sum_{i=1}^n \hat{p}_{ig}.$$

M-askel: Maksimoidaan logaritmoitu uskottavuusfunktio (4.24) mallin jokaiselle komponentille erikseen käyttäen painona posterioritodennäköisyyksiä

$$\max_{\theta_g} \sum_{i=1}^n \hat{p}_{ig} \log f(y_i|x_i, \theta_g).$$

E- ja M-askeleita toistetaan niin kauan, kunnes uskottavuuden paraneminen jää alle ennalta määritetyn kynnyksen tai kunnes maksimimäärä iteraatioita on saavutettu.

4.3.1 Akaiken informaatiokriteeri

Mallin komponenttien $g=1, 2, \dots, G$ määrän valinnassa käytetään apuna Akaiken informaatiokriteeriä:

$$(4.24) \quad \text{AIC} = -2(\log(L(\hat{\psi})) + 2d),$$

missä L on uskottavuusfunktio ja d mallissa olevien parametrien lukumäärä. Eri määrän komponentteja sisältäviä malleja voidaan vertailla Akaiken informaatiokriteerin perusteella. Mitä pienempi on mallin AIC, sitä parempi malli on kyseessä. Mallien vertailuun on olemassa myös muita menetelmiä, mutta tässä tutkielmassa käytetään tätä.

5 Aineiston analyysi

5.1 Painoindeksikäyrien mallinnus

Tässä tutkielmassa lasten painoindeksikäyriä mallinnettiin kahdella eri tavalla: neljännen asteen polynomifunktioilla sekä kuutiotasoitetuilla splineilla.

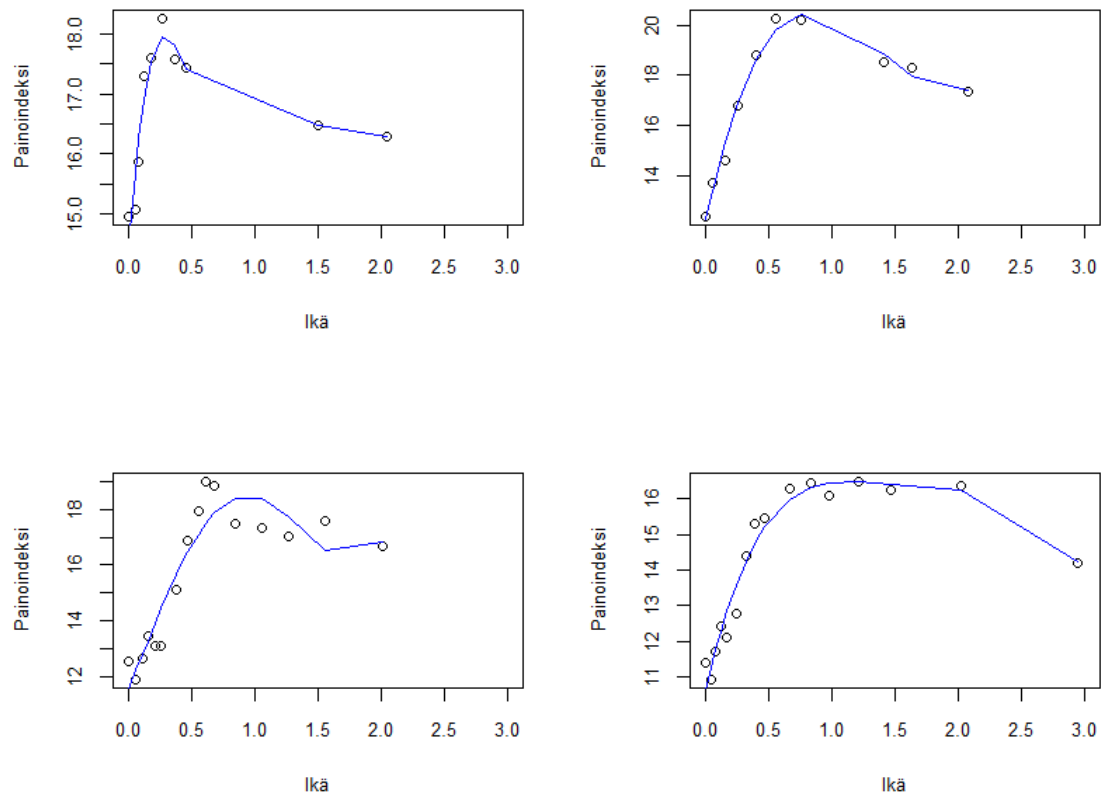
5.1.1 Painoindeksin varhaislapsuuden piikin mallinnus

Lasten painoindeksejä syntymän ja kolmen ikävuoden välillä mallinnettiin neljännen asteen polynomiregressiomallein. Mallinnustavan valintaa varten aineistosta poimittiin satunnaisesti 12 lapsen testijoukko, joille mallinnettiin painoindeksit iän suhteen. Visuaalisen tarkastelun avulla tutkittiin mallinnuksen sopivuutta. Kuutiotasoitettut splinit osoittautuivat tässä tapauksessa epäsopiviksi, sillä lapsilla oli tällä ikävälillä paljon mittapisteitä ja painoindeksit vaihtelivat mittapisteiden välillä kumpaankin suuntaan. Tämän seurauksena sekä ristiinvalidointi että yleistetty ristiinvalidointi valitsivat liian pienen tasoitusparametrin arvon ja splinikäyristä olisi tullut liian tarkkoja eli tässä tapauksessa liian mutkittelevia. Painoindeksin varhaislapsuuden piikin määrittäminen näiltä käyriltä olisi ollut vaikeaa. Silmämääräinen tasoitusparametrin valinta ei tullut kyseeseen, sillä lasten mittausten lukumäärät vaihtelivat paljon, jolloin ei voida määrittää yhtä, koko aineistolle sopivaa tasoitusparametrin arvoa.

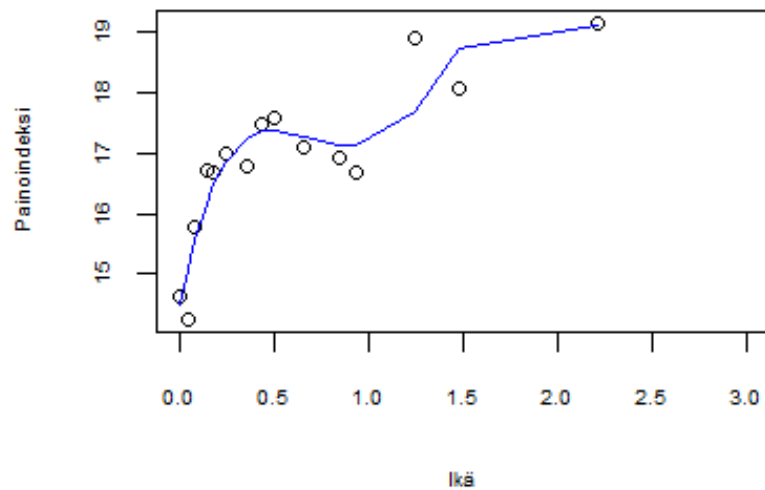
Mallinnustavaksi valikoitu neljännen asteen polynomiregressio, sillä mallinnus tällä ikävälillä sai olla hieman karkeampi. 12 lapsen testijoukolle sovitettiin kolmannen, neljännen ja viidennen asteen polynomiregressiomallit ja vertailtiin niiden sopivuutta sekä silmämääräisesti että varianssianalyysin avulla. Neljännen asteen polynomit sopivat yleisesti parhaiten kaikille testijoukon lapsille.

Mallinnuksen sopivuuden tarkastelun jälkeen tehtiin mallinnus koko aineistolle. Sovitettiin kaikille lapsille neljännen asteen polynominen regressiomalli, missä painoindeksi oli selitettävänä muuttujana ja selittäjänä oli iälle sovitettu neljännen asteen polynomi. Lapsen painoindeksin varhaislapsuuden piikki määritettiin olevan neljännen asteen polynomisen regressiomallin suurin estimaatti välillä $0 < x < 1.5$. Välin ääripää oli siis suljettu pois vaihtoehdoista. Estimaattien avulla pystyttiin muodostamaan painoindeksikäyrät iän suhteen. Kuviossa 5.1 on esitetty neljän lapsen painoindeksille iän suhteen mallinnetut neljännen asteen polynomiregressiokäyrät. Lisäksi kuviossa 5.2 on esitetty yhden lapsen painoindeksikäyrä, josta painoindeksin varhaislapsuuden piikkiä ei pystytty määrittämään.

Painoindeksin varhaislapsuuden piikki pystyttiin mallintamaan 1061:lle lapselle. Estimoitu painoindeksin piikki oli keskimäärin 0.67 vuoden iässä (8 kk), tytöillä 0.7-vuotiaana (8.4 kk) ja pojilla jo aikaisemmin 0.65-vuotiaana (7.8 kk) ($p < 0.001$). Kuviossa 5.3 on esitetty painoindeksin varhaislapsuuden piikin ajankohdan jakaumat sukupuolittain. Kuvioista nähdään, että jakaumat eroavat jonkin verran toisistaan ja ovat molemmat hieman oikealle vinoutuneita. Painoindeksin varhaislapsuuden piikin ajankohtien mediaani oli 0.65 vuotta (7.8 kk) ja sukupuolittain tytöillä



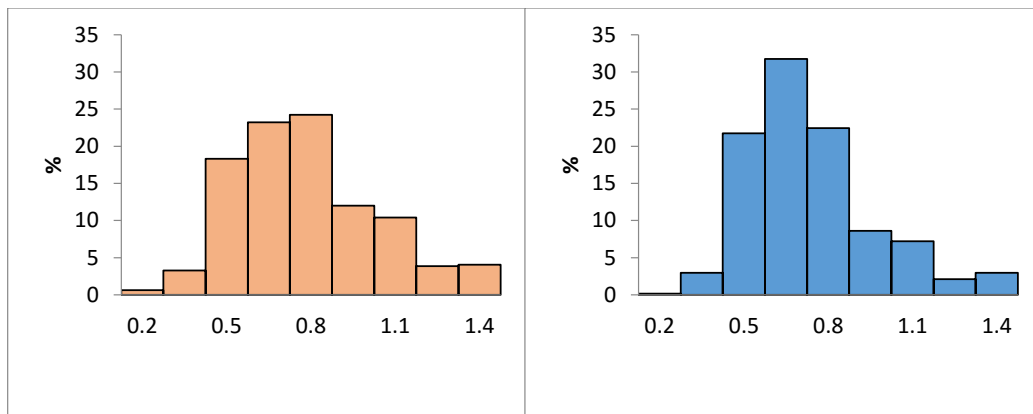
Kuvio 5.1. Neljän lapsen neljännen asteen polynomien avulla iän suhteen mallinnetut painoindeksikäyrät. Painoindeksin varhaislapsuuden piikit ovat järjestyksessä ylärivin vasemmasta kuviosta alkaen 0.27, 0.75, 1.05 ja 1.21 vuoden iässä ja estimoidut painoindeksit näissä kohdissa 18.3, 20.2, 17.4 ja 16.5 kg/m^2 .



Kuvio 5.2. Yhden lapsen neljännen asteen polynomien avulla iän suhteen mallinnettu painoindeksikäyrä, josta painoindeksin varhaislapsuuden piikkiä ei voitu määrittää.

0.67 vuotta (8 kk) ja pojilla 0.59 vuotta (7.1 kk). Aiemmassa kirjallisuudessa painoindeksin varhaislapsuuden piikin ajankohtien estimoidut keskiarvot ja mediaanit vaihtelevat seitsemän ja lähes kymmenen kuukauden iän välillä. Tässä tutkielmassa estimoitu painoindeksin piikin ajankohta osuu myös tälle ikävälille. Useimmissa tutkimuksissa painoindeksin piikki oli pojilla aikaisemmin kuin tytöillä (Silverwood et al. (2009), Mook-Kanamori et al. (2011) ja Hof et al. (2013)), niin kuin se oli myös tässä tutkielmassa. Painoindeksin piikkiä ei ole vielä paljon tutkittu, joten siihen liittyvää kirjallisuutta on saatavilla melko vähän.

Estimoitu painoindeksi varhaislapsuuden piikin kohdalla oli keskimäärin 17.7 kg/m^2 , tytöillä 17.4 kg/m^2 ja pojilla 17.9 kg/m^2 ($p < 0.001$). Nämä ovat hyvin samaa luokkaa kuin aiemmassa kirjallisuudessa, missä BMI painoindeksin piikissä on keskimäärin estimoitu olevan $17\text{--}18 \text{ kg/m}^2$ välillä ja pojilla korkeampi kuin tytöillä. Tyttöjen ja poikien varhaislapsuuden piikin painoindeksit noudattivat melko hyvin normaalijakaumaa, vaikkakin molemmat jakaumat olivat hieman oikealle vinoutuneita.



Kuvio 5.3. Vasemmalla tyttöjen ja oikealla poikien painoindeksin varhaislapsuuden piikin ajankohdan (vuosina) jakauma.

Polynomiset regressiomallit estimoitiin R-ohjelmiston "lm"-funktioilla, johon selittävälle x -muuttujalle sovitettiin kolmannen, neljännen ja viidennen asteen polynomi käyttäen "poly"-funktioita. Eri asteisia polynomiregressiomalleja vertailtiin "anova"-funktion avulla. Mallin parametrien estimaatit saatiin laskettua "predict"-funktion avulla. Malli sovitettiin välille $0 \leq x < 3$ ja estimaateista poimittiin maksimiarvo väliltä $0 < x < 1.5$. Lisäksi funktiolle määriteltiin, että lapsella täytyi olla vähintään kuusi mittapistettä ikävälillä, johon malli sovitettiin.

5.1.2 Painoindeksin käännepisteen mallinnus

Lasten painoindeksejä kahden ja kymmenen ikävuoden välillä mallinnettiin kuutio-tasoitetuilla splineilla, joihin tasoitusparametri valittiin yleistetyn ristiinvalidoinnin avulla. Mallinnustavan valintaa varten aineistosta poimittiin satunnaisesti 12 lapsen testijoukko, jolle mallinnettiin painoindeksit iän suhteen. Visuaalisen tarkastelun

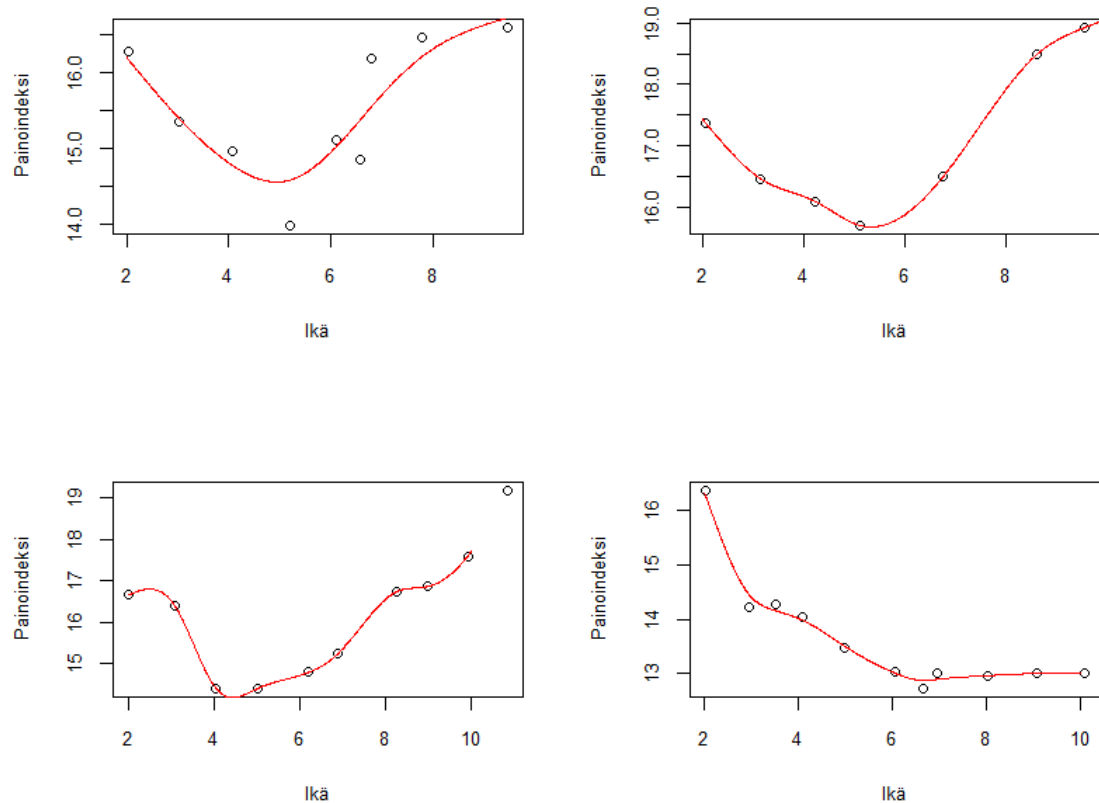
avulla tutkittiin mallinnuksen sopivuutta sekä sopivaa tasoitusparametrin valintaan käytettävää menetelmää. Kuutiotasoitettujen splinien mallinsivat hyvin painoindeksien ja painoindeksien käännepesteet päätettiin estimoida etsimällä splinikäyrien minimiarvot kahden ja kymmenen ikävuoden välillä. Perusteena tälle estimointitavalle oli se, että käyrän minimikohdan jälkeen painoindeksissä täytyy olla nouseva trendi, koska käyrä ei tämän jälkeen saavuta enää pienempää arvoa kuin minimiarvo. Koska painoindeksissä on nouseva trendi, käännepesteen voidaan ajatella sijaitsevan tässä minimikohdassa. Tasoitusparametrin valintaan parhaaksi menetelmäksi osoittautui yleistetty ristiinvalidointi. Tavallinen ristiinvalidointi valitsi monessa tapauksessa liian suuren tasoitusparametrin arvon, jolloin mallinnettu splinikäyrä oli turhan karkea. Silmämääräinen valinta ei tullut kyseeseen, sillä lasten mittausten lukumäärät vaihtelivat paljon, jolloin ei voida määrittää yhtä, koko aineistolle sopivaa tasoitusparametrin arvoa.

Mallinnuksen sopivuuden tarkastelun jälkeen tehtiin mallinnus koko aineistolle. Spliniestimaateille laskettiin arvot 0.01:n ikävuoden välein, koska aineistossa lasten iät on ilmoitettu kahden desimaalin tarkkuudella. Lapsen painoindeksien käännepeste määritettiin olevan pienin estimaatti välillä $2 < x < 10$. Estimattien avulla pystyttiin muodostamaan painoindeksikäyrät iän suhteen. Kuviossa 5.4 on esitetty iän suhteen mallinnetut painoindeksikäyrät samoille lapsille vastaavassa järjestyksessä kuin kuviossa 5.1 painoindeksien varhaislapsuuden piikin mallinnuksen yhteydessä. Lisäksi kuviossa 5.5 on esitetty yhdelle lapselle iän suhteen mallinnettu painoindeksikäyrä, josta painoindeksien käännepesteä ei pystytty määrittämään. Käyrässä ei ole havaittavissa käännekohtaa, jonka jälkeen painoindeksi lähtisi nousuun.

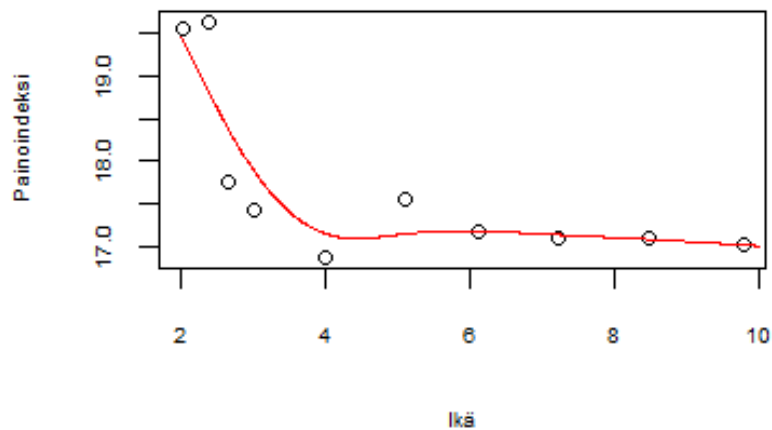
Lasten painoindeksien käännepeste oli keskimäärin 5.6 vuoden iässä, tytöillä 5.5-vuotiaana ja pojilla 5.7-vuotiaana. Tyttöjen ja poikien ero käännepesteen ajankohdassa ei ollut tilastollisesti merkitsevä. Nämä täsmäivät hyvin aiempaan kirjallisuuteen, missä käännepesteen ollessa jatkuva muuttuja, se on keskimäärin ollut 5.5-6 ikävuoden tienoilla ja pojilla hieman myöhemmin tai lähes samaan aikaan kuin tytöillä. Kuviossa 5.6 on esitetty painoindeksien käännepesteen ajankohdan jakaumat sukupuolittain. Kuvioista nähdään, että vaikka tyttöjen ja poikien käännepesteen ajankohdat olivat keskimäärin lähes samaan aikaan, niiden jakaumat erosivat jonkin verran toisistaan. Molemmat jakaumat noudattavat hyvin normaalijakaumaa, mutta tytöillä jakauma näyttäisi painottuvan enemmän aikaisempiin ajankohtiin kuin pojilla ja päinvastoin pojilla enemmän myöhempiin ajankohtiin kuin tytöillä. Tämän vuoksi on perusteltua rakentaa mallit myös erikseen kummallekin sukupuolelle.

Estimoitu painoindeksi käännepesteessä oli keskimäärin 15.2 kg/m^2 sekä tytöillä että pojilla. Useassa painoindeksien käännepesteeseen liittyvässä tutkimuksessa on määritetty vain käännepesteen ajankohta eikä lainkaan painoindeksiä tässä kohtaa. Niissä, joissa on määritetty myös painoindeksi, se on ollut $15.5\text{--}16 \text{ kg/m}^2$ paikkeilla. Tässä tutkielmassa estimoitu käännepesteen BMI on melko samaa luokkaa, mutta hieman pienempi. Tyttöjen ja poikien painoindeksit olivat suunnilleen normaalisti jakautuneita, hieman oikealle vinoutuneita.

Splini-mallit estimointiin R-ohjelmiston "smooth.spline"-funktioilla, johon on mahdollista määrittellä tasoitusparametri itse tai valita sen valintamenetelmäksi ristiinvalidointi tai yleistetty ristiinvalidointi. Oletuksena funktiossa on ristiinvalidointi,

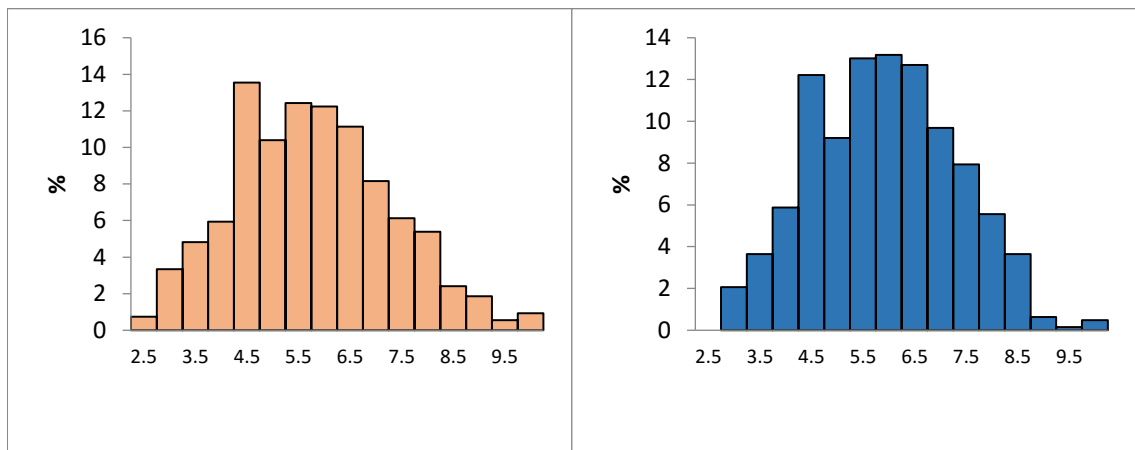


Kuvio 5.4. Neljän lapsen kuutiotasoitettujen splinien avulla iän suhteen mallinnetut painoindeksikäyrät. Painoindeksin käännepestteet ovat järjestyksessä ylä-rivin vasemmasta kuviosta alkaen 4.95, 5.34, 4.43 ja 6.66 vuoden iässä ja estimoidut painoindeksit näissä kohdissa 14.6, 15.7, 14.2 ja 12.9 kg/m².



Kuvio 5.5. Yhden lapsen kuutiotasoitettujen splinien avulla iän suhteen mallinnettu painoindeksikäyrä, josta painoindeksin käännepestettä ei voitu määrittää.

joten valintamenetelmäksi määriteltiin erikseen yleistetty ristiinvalidointi. Tämä tapahtui komennolla "cv=FALSE". Estimaatit saatiin laskettua "predict"funktion avulla.



Kuvio 5.6. Vasemmalla tyttöjen ja oikealla poikien painoindeksin käänne-
steen ajankohdan (vuosina) jakauma.

la. "Predict"-funktiolle määriteltiin "seq"-komennon avulla, että estimaatit lasketaan väliltä $2 \leq x \leq 10$ 0.01:n ikävuoden välein. Estimaateista poimittiin pienin väliltä $2 < x < 10$ eli välin ääripäät suljettiin pois vaihtoehdoista. Jos minimiarvoa ei löytynyt, ei käännepestettä pystytty määrittämään kyseiselle lapselle. Lisäksi funktiolle määriteltiin, että lapsella täytyi olla vähintään kuusi mittapistettä, joista ainakin yksi on oltava viiden ikävuoden jälkeen.

5.2 Lapsen painoindeksin käännepesteen ajankohtaan vaikuttavat tekijät

Tutkittiin ensin lapsen käännepesteen ajankohtaan mahdollisesti vaikuttavia tekijöitä yksinkertaisten lineaaristen mallien avulla, jotta voidaan nähdä, osoittautuuko joku tutkittavista tekijöistä olevan yhteydessä käännepesteen ajankohtaan. Tärkeimmät tutkittavat tekijät olivat äidin ennen raskautta mitattu painoindeksi, raskauden loppuvaiheessa mitattu painoindeksi, painoindeksin muutos raskauden aikana, diabetes, pituus ja paino sekä lapsen syntymämittaukset ja varhaislapsuuden painoindeksin piikin ajankohta sekä painoindeksi tässä kohtaa. Myös muille saatavilla oleville muuttujille muodostettiin yksinkertaiset lineaariset regressiomallit. Tämän jälkeen muodostettiin useamman selittäjän lineaarisia regressiomalleja, joihin valittiin sopivat muuttujayhdistelmät. Kaikki mallit tehtiin koko aineistolle sekä myös erikseen tytöille ja pojille.

5.2.1 Yksinkertaiset lineaariset regressiomallit

Taulukossa 5.1 on esitetty yksinkertaisten lineaaristen mallien tulokset sekä koko aineistolle että tytöille ja pojille erikseen tehdyistä malleista. Taulukossa jokaisella rivillä on yksi lineaarinen regressiomalli ja tässä mallissa käytetty selittävä muuttuja

on mainittu rivin alussa. Poikkeuksena diabetes, joka on kolmiluokkainen kvalitatiivinen muuttuja ja on näin ollen mallissaan kahtena dummy-muuttujana. Diabeteksen vertailuryhmänä on käytetty diabetesta sairastamattomia ja mallissa kahta muuta ryhmää on verrattu tähän ryhmään. Diabetes-muuttujan malli on tämän vuoksi taulukossa kahdella rivillä. Sarakkeilla on jokaisen mallin vakiotermien estimaatti, parametrin estimaatti, p-arvo, mallin residuaalien keskivirhe sekä mallin korjattu selitysaste. Yhden selittäjän lineaarisissa malleissa selittävän muuttujan p-arvo on sama kuin koko mallin p-arvo, lukuunottamatta diabeteksen mallia, joka sisältää p-arvot erikseen kummallekin ryhmävertailulle. Tämän koko mallin p-arvo oli <0.001.

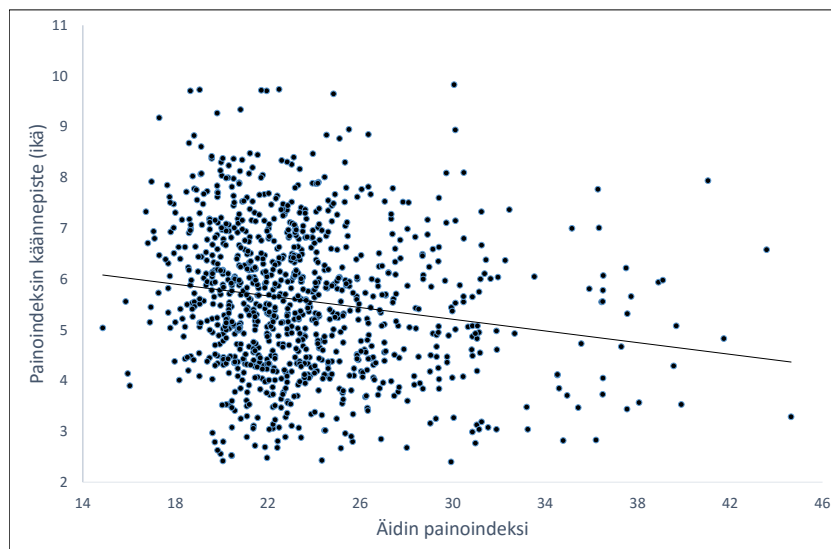
Taulukko 5.1. Yksinkertaisten lineaaristen regressiomallien tulokset koko aineistolle sekä erikseen tytöille ja pojille.

	Kaikki					Tytöt					Pojat				
	Vakiotermien estimaatti	Parametrin estimaatti	P-arvo	Residuaalien keskivirhe	Korjattu selitysaste	Vakiotermien estimaatti	Parametrin estimaatti	P-arvo	Residuaalien keskivirhe	Korjattu selitysaste	Vakiotermien estimaatti	Parametrin estimaatti	P-arvo	Residuaalien keskivirhe	Korjattu selitysaste
Äidin BMI	6.94	-0.06	<0.001	1.41	2.7 %	6.60	-0.05	0.004	1.47	1.5 %	7.21	-0.07	<0.001	1.35	3.8 %
Äidin paino	6.84	-0.02	<0.001	1.41	2.7 %	6.45	-0.01	0.008	1.47	1.3 %	7.17	-0.02	<0.001	1.35	4.1 %
Äidin pituus	6.38	0.00	0.520	1.43	-0.1 %	4.87	0.00	0.742	1.48	-0.2 %	7.70	-0.01	0.209	1.37	0.1 %
Diabetes - raskauden aikainen	5.69	-0.42	<0.001	1.44	1.7 %	5.61	-0.41	0.021	1.52	1.1 %	5.76	-0.43	0.003	1.38	1.9 %
Diabetes - tyyppi I	5.69	-1.07	0.003	1.44	1.7 %	5.61	-1.12	0.074	1.52	1.1 %	5.76	-1.07	0.011	1.38	1.9 %
Äidin BMI (raskaus)	7.49	-0.07	<0.001	1.41	3.3 %	6.96	-0.05	0.060	1.46	1.4 %	7.88	-0.08	<0.001	1.36	4.7 %
BMI muutos	5.67	-0.02	0.694	1.44	-0.2 %	5.66	-0.04	0.525	1.47	-0.3 %	5.75	-0.01	0.863	1.40	-0.4 %
Syntymäpaino	5.78	-0.05	0.489	1.45	0.0 %	5.48	0.00	0.917	1.53	-0.2 %	6.11	-0.12	0.194	1.39	0.1 %
Syntymäpituus	5.55	0.00	0.951	1.46	-0.1 %	5.17	0.01	0.804	1.53	-0.2 %	6.25	-0.01	0.612	1.39	-0.1 %
Syntymä BMI	5.90	-0.02	0.453	1.46	0.0 %	5.26	0.02	0.691	1.53	-0.2 %	6.41	-0.05	0.143	1.39	0.2 %
Syntymä- pään ympäry	6.84	0.04	0.255	1.42	0.0 %	7.84	-0.07	0.206	1.47	0.2 %	6.99	-0.04	0.346	1.38	0.0 %
Painoindeksi piikin ajankohta	5.59	0.04	0.828	1.47	-0.1 %	5.19	0.47	0.093	1.54	0.4 %	5.92	-0.34	0.196	1.39	0.1 %
Painoindeksi piikin BMI	5.48	0.01	0.792	1.47	-0.1 %	5.34	0.01	0.809	1.54	-0.2 %	5.86	-0.01	0.791	1.40	-0.2 %

Taulukosta 5.1 nähdään, että lapsen painoindeksiin käännepestettä tilastollisesti merkitsevästi selittäviä muuttujia ovat äidin painoindeksi ennen raskautta sekä raskaudenseurannan viimeisessä mittauksessa, äidin paino ja diabetes. Tehtiin myös malli, jossa painoindeksiin käännepesteen aikainen painoindeksi oli selittävänä muuttujana käännepesteen ajankohdalle. Myös tämä malli oli tilastollisesti erittäin merkitsevä. Tyttöjen ja poikien malleissa tulokset hieman eroavat toisistaan. Tytöillä äidin viimeisen raskaudenseurantamittauksen painoindeksi ei ollut mallissaan tilastollisesti merkitsevä. Myös muut tilastollisesti merkitsevät mallit olivat vähemmän merkitseviä tytöillä kuin pojilla. Missään mallissa korjattu selitysaste ei ollut erityisen korkea, tytöillä se oli monesti vielä matalampi kuin pojilla. Korjatut selity-

sasteet olivat yhden ja viiden prosentin välillä tilastollisesti merkitsevissä malleissa. Poikkeuksena mallissa, missä painoindeksin käänne pisteen ajankohta oli selittäjänä, korjattu selitysaste oli hieman korkeampi, pojilla jopa 9.1 %. Tämä tarkoittaa käytännössä sitä, että mikään näistä tutkittavista muuttujista ei selitä kovin hyvin lapsen painoindeksin käänne pisteen ajankohdan vaihtelua. Muista tehdyistä malleista ei löydetty tilastollisesti merkitseviä selittäjiä lapsen käänne pisteen ajankohdalle.

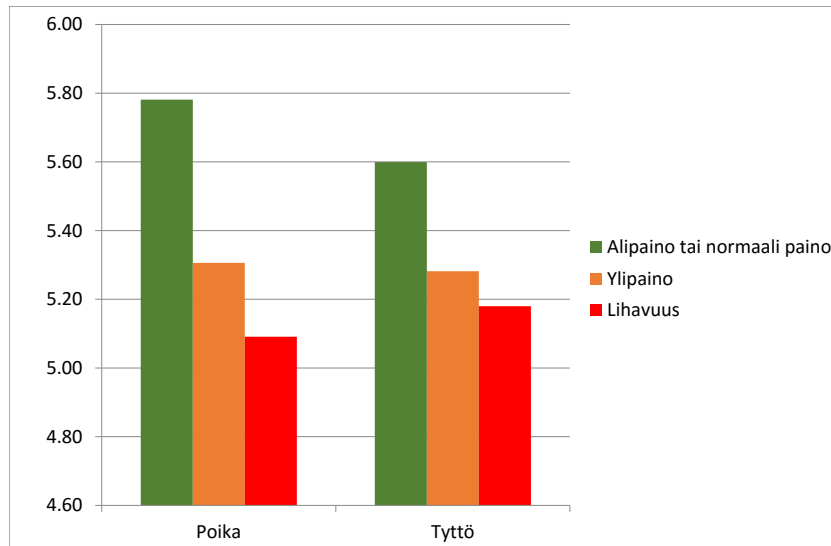
Äidin painoindeksi näyttäisi olevan yhteydessä lapsen painoindeksin käänne pisteseen. Tämä tulos vahvistaa aiempien tutkimusten havaintoja aiheesta. Äidin painoindeksin vakiotermi estimaatti oli 6.94 ja parametrin estimaatti -0.06. Tämä tarkoittaa teoriassa sitä, että äidin painoindeksin ollessa 0, lapsen painoindeksin käänne pisteen ajankohta olisi 6.94 vuotiaana ja tästä lähtien aina kun äidin painoindeksi on yhtä yksikköä suurempi, lapsen painoindeksin käänne pisteen ajankohta on keskimäärin 0.06 vuotta eli hieman vajaa yhden kuukauden aikaisemmin. Toisin sanoen siis äidin suuremmalla painoindeksillä näyttäisi tämän perusteella olevan yhteys lapsen aikaistuneeseen painoindeksin käänne pisteen ajankohtaan. Äidin painoindeksi selittää kuitenkin vain 2.7 % lapsen käänne pisteen ajankohdan vaihtelusta eli vaikka yhteys oli tilastollisesti merkitsevä, ei äidin painoindeksin selitysvoima ole kovin suuri. Yhteyttä havainnollistaa kuvio 5.7, missä on esitetty äidin painoindeksin ja lapsen painoindeksin käänne pisteen ajankohdan välinen korrelaatiodiagrammi ja siihen sovitettu lineaarinen regressiosuora. Kuvioista nähdään näiden muuttujien välinen negatiivinen korrelaatio.



Kuvio 5.7. Äidin painoindeksin ja lapsen painoindeksin käänne pisteen välinen korrelaatiodiagrammi ja siihen sovitettu lineaarinen regressiosuora.

Tyttöjen ja poikien malleissa estimaatit hieman erosivat toisistaan. Pojilla äidin korkeamman painoindeksin käänne pistettä aikaistava vaikutus oli hieman suurem-

paa kuin tytöillä. Myös korjattu selitysaste oli poikien mallissa selvästi suurempi kuin tytöillä. Kuviossa 5.8 on esitetty tyttöjen ja poikien käännepisteen keskimääräiset ajankohdat äidin luokitellun painoindeksin mukaan. Myös siitä käy ilmi, että pojilla äidin painoindeksin yhteys käännepisteen ajankohtaan näyttäisi olevan vahvempi kuin tytöillä.



Kuvio 5.8. Tyttöjen ja poikien käännepisteen keskimääräiset ajankohdat äidin luokitellun painoindeksin mukaan.

Äidin paino ennen raskautta oli myös malleissaan merkitsevä selittäjä ja sen vaikutus lapsen käännepisteeseen oli samansuuntaista, mutta vähäisempää kuin painoindeksin. Äidin raskaudenseurannan viimeisen mittauksen painoindeksin vaikutus käännepisteen ajankohtaan oli hyvin samanlaista kuin äidin ennen raskautta mitatun painoindeksin. Raskauden aikainen painoindeksin tai painon muutos (ei esitetty taulukossa) eivät olleet tilastollisesti merkitseviä selittäjiä käännepisteen ajankohdalle.

Äidin diabetes osoittautui myös olevan yhteydessä lapsen käännepisteen ajankohtaan. Mallien korjatut selitysasteet olivat kuitenkin vain 1-2 prosenttia luokkaa. Kolmiluokkainen diabetes-muuttuja oli malleissa kahtena dummy-muuttujana. Vertailuryhmänä malleissa oli diabetesta sairastamattomat äidit. Näihin verrattuna raskauden aikaista tai tyypin I diabetesta sairastavien äitien lasten painoindeksin käännepiste oli ollut aikaisemmin ja erot olivat tilastollisesti merkitseviä lukuunottamatta tyttöjen mallissa tyypin I diabeetikkojen ja diabetesta sairastamattomien äitien lasten painoindeksin käännepisteen välistä eroa. Tyypin I diabeetikkoja aineistossa oli kuitenkin hyvin vähän niin kuin on jo aikaisemmin todettu, joten ei ole kovin mielekästä edes käyttää vertailussa tyypin I diabeetikkoja. Taulukossa 5.2 on esitetty lapsen käännepisteen ajankohdan ryhmittäiset keskiarvot, keskihajonnat sekä

havaintojen lukumäärät diabetes-muuttujan mukaan ryhmiteltynä. Tunnusluvut on esitetty myös erikseen tytöille ja pojille.

Taulukko 5.2. Lapsen painoindeksin käännepisteen ajankohdan tunnuslukuja äidin diabeteksen mukaan koko aineistolle sekä tytöille ja pojille erikseen.

	Tytöt	Pojat	Kaikki
Ei	5.61	5.76	5.69
diabetestä,	(1.50)	(1.35)	(1.42)
KA(sd)	N=443	N=508	N=951
Raskauden			
aikainen	5.2	5.33	5.27
diabetes,	(1.64)	(1.44)	(1.53)
KA(sd)	N=88	N=110	N=198
Tyypin I	4.49	4.7	4.62
diabetes,	(0.88)	(1.81)	(1.52)
KA(sd)	N=6	N=11	N=17

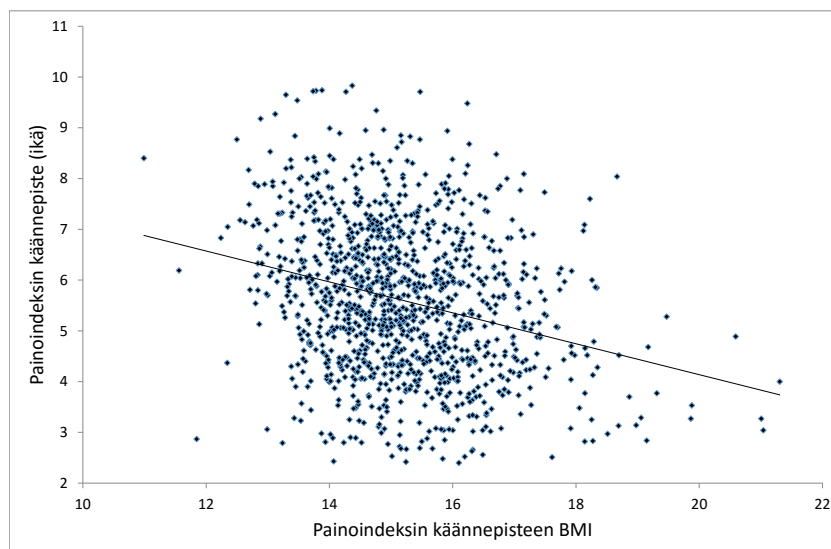
Lasten syntymämittauksista mikään ei ollut merkitsevä selittäjä mallissaan. Tyttöillä syntymäpainon, -pituuden ja -painoindeksin parametrien estimaatit olivat positiivisia kun taas pojilla ne olivat negatiivisia. Mallissa, jossa selittäjänä oli lapsen päänympärys syntyessä, parametrien estimaatit olivat negatiivisia molemmilla sukupuolilla. Taulukossa 5.3 on esitetty lapsen käännepisteen keskimääräiset ajankohdat syntymäpainon mukaan (erittäin pieni, normaali, suuri, erittäin suuri) sekä koko aineistolle että tytöille ja pojille erikseen. Lisäksi suluissa on kerrottu keskimääräinen painoindeksi käännepisteessä ja alimpana havaintojen lukumäärät. Aikaisemmissa tutkimuksissa oli havaittu yhteys suuren syntymäpainon ja aikaistuneen käännepisteen välille. Tässä tukielmassa yksinkertaisten lineaarisien mallien perusteella eri painoluokkien välillä ei havaittu tilastollisesti merkitseviä eroja painoindeksin käännepisteen ajankohdissa. Tyttöillä kaikista aikaisin käännepiste on ollut niillä, joiden syntymäpaino oli ollut pieni, kun taas pienenä syntyneillä pojilla käännepiste on ollut kaikista myöhäisin kun verrataan suurempiin syntymäpainoluokkiin. Malleissa vertailtiin painoluokkia alle 2.5 kg, 2.5-4 kg ja yli 4 kg.

Lapsen painoindeksin varhaislapsuuden piikin ajankohta tai painoindeksi eivät myöskään selittäneet lapsen painoindeksin käännepisteen ajankohtaa tilastollisesti merkitsevästi. Ainoastaan tytöillä painoindeksin piikin ajankohdan positiivinen parametrien estimaatti oli melkein tilastollisesti merkitsevä (estimaatti 0.47 ja $p=0.093$). Poikien vastaava estimaatti oli negatiivinen ja tilastollisesti ei-merkitsevä (estimaatti -0.34 ja $p=0.196$). Aikaisemmissa tutkimuksissa varhaislapsuuden piikin myöhäisemmän ajankohdan sekä suuremman BMI:n oli todettu olevan yhteydessä suurempaan painoindeksiin myöhemmässä lapsuudessa (Silverwood et al. (2009), Mook-Kanamori et al. (2011), Hof et al. (2013) ja Wen et al. (2012)). Tästä voisi ajatella, että nämä tekijät saattaisivat olla yhteydessä myös aikaistuneeseen painoindeksin käännepisteeseen. Ainakin yhdessä tutkimuksessa oli löydettykin heikko negatiivinen korrelaatio (-0.09) piikin ajankohdan ja käännepisteen ajankohdan välille (Wen

Taulukko 5.3. Lapsen painoindeksin käännepisteen ajankohdan ja käännepisteen BMI:n (suluissa) keskiarvot syntymäpainon mukaan koko aineistolle sekä tytöille ja pojille erikseen.

Syntymäpaino	Tytöt	Pojat	Kaikki
	5.11	5.79	5.52
	(14.60)	(14.95)	(14.81)
< 2.5 kg	N=24	N=36	N=60
	5.55	5.68	5.62
	(15.60)	(15.07)	(15.10)
2.5 - 4 kg	N=452	N=473	N=925
	5.47	5.59	5.55
	(16.01)	(15.86)	(15.69)
> 4kg	N=62	N=121	N=183
	5.01	5.87	5.67
	(15.13)	(15.74)	(15.90)
> 4.5 kg	N=6	N=20	N=26

et al. (2012)). On mahdollista, että myös tässä tutkielmassa yhteyksiä löytyy, kun malliin laitetaan useampi selittävä muuttuja.



Kuvio 5.9. Painoindeksin käännepisteen BMI:n yhteys painoindeksin käännepisteen ajankohtaan.

Tehtiin vielä malli, missä lapsen käännepisteen ajankohtaa selitti lapsen painoindeksi käännepisteessä. Malli oli tilastollisesti erittäin merkitsevä ja painoindeksi käännepisteessä osoittautui kaikista parhaaksi selittäjäksi tässä aineistossa. Sen

parametrin estimaatti oli -0.30 eli tämän perusteella näyttäisi, että aikaista kään-
nepistettä edeltäisi suurempi painoindeksi kuin myöhäisempää käännepistettä. Tätä
yhteyttä havainnollistaa kuvio 5.9.

5.2.2 Useamman selittäjän lineaariset regressiomallit

Muodostettiin seuraavaksi useamman selittäjän lineaarisia regressiomalleja. Mallei-
hin sisällytettiin muuttujat, joiden yhteyttä käännepisteeseen haluttiin tutkia. Lisäksi
malleihin sisällytettiin äidin raskauden kesto, äidin ikä sekä äidin aiempien synny-
tysten lukumäärä, jotta nähdään, vaikuttavatko nämä taustamuuttujat mallin tulok-
siin. Äidin kaksi painomittausta sekä kaksi painoindeksimittausta korreloivat kaikki
niin vahvasti keskenään, että niitä ei ole järkevää laittaa samaan malliin. Myös lap-
sen syntymämittauksista painoindeksi jätettiin malleista pois, sillä se on muodos-
tettu lapsen pituuden ja painon perusteella, jotka on jo sisällytetty malleihin. Mallit
tehtiin erikseen tytöille ja pojille.

Ensimmäinen useamman selittäjän lineaarinen regressiomalli sisälsi seuraavat
selittävät muuttujat: äidin painoindeksi ennen raskautta, äidin diabetes, äidin pituus,
lapsen syntymäpaino, lapsen syntymäpituus, lapsen päänympärysy syntyessä, lapsen
painoindeksiin piikin ajankohta, lapsen painoindeksiin piikin BMI, raskauden kesto,
äidin ikä sekä äidin aiempien synnytysten lukumäärä. Malli tehtiin erikseen tytöille
ja pojille. Tyttöjen mallin tulokset on esitetty taulukossa 5.4 ja poikien malli tulokset
taulukossa 5.5.

Taulukko 5.4. Tyttöjen useamman selittäjän lineaarisen regressiomallin tulok-
set. Mallin AIC=1272.0.

	Estimaatti	Keskivirhe	T-testisuure	P-arvo
Vakio	6.92	4.05	1.71	0.088
Äidin painoindeksi	-0.04	0.02	-1.67	0.095
Diabetes - Raskauden aikainen diabetes	-0.04	0.24	-0.17	0.868
Diabetes - Tyypin I diabetes	-0.43	0.76	-0.57	0.570
Äidin pituus	0.00	0.01	0.06	0.956
Syntymäpaino	-0.44	0.29	-1.50	0.136
Syntymäpituus	0.05	0.06	0.92	0.359
Päännympärysy	0.08	0.08	0.98	0.329
Piikin ajankohta	0.47	0.32	1.45	0.148
Piikin BMI	-0.05	0.05	-0.93	0.352
Raskaudenkesto (pv)	-0.01	0.01	-1.45	0.147
Äidin ikä (pv)	0.00	0.00	-1.24	0.215
Aiemmat synnytykset	0.08	0.09	0.89	0.376
Residuaalien keskivirhe	1.44			
Selitysaste	4.0 %			
Korjattu selitysaste	0.6 %			
F-testisuure	1.19			
P-arvo	0.292			
Vapausasteet	339			

Taulukko 5.5. Poikien useamman selittäjän lineaarisen regressiomallin tulokset. Mallin AIC=1443.4.

	Estimaatti	Keskivirhe	T-testisuure	P-arvo
Vakio	8.71	3.44	2.53	0.012
Äidin painoindeksi	-0.07	0.02	-3.44	0.001
Diabetes - Raskauden aikainen diabetes	-0.06	0.22	-0.27	0.784
Diabetes - Tyypin I diabetes	-0.19	0.59	-0.32	0.750
Äidin pituus	-0.02	0.01	-1.60	0.109
Syntymäpaino	-0.31	0.24	-1.29	0.198
Syntymäpituus	0.03	0.05	0.64	0.521
Päänympärys	0.04	0.06	0.65	0.514
Piikin ajankohta	-0.19	0.31	-0.61	0.540
Piikin BMI	0.00	0.04	-0.07	0.946
Raskaudenkesto (pv)	0.00	0.01	-0.03	0.976
Äidin ikä (pv)	0.00	0.00	-0.29	0.769
Aiemmat synnytykset	0.11	0.08	1.46	0.146
Residuaalien keskivirhe	1.36			
Selitysaste	5.6 %			
Korjattu selitysaste	2.7 %			
F-testisuure	1.96			
P-arvo	0.026			
Vapausasteet	400			

Tyttöjen mallissa merkitseviä selittäjiä käännepesteen ajankohdalle ei ole havaittavissa. Ainoastaan äidin painoindeksi on melkein merkitsevä (estimaatti -0.04 ja $p=0.095$). Myöskään mallissa oleva muuttujayhdistelmä ei selitä painoindeksin käännepesteen ajankohdan vaihtelua merkitsevästi, sillä kokonaisuudessaan malli ei ole tilastollisesti merkitsevä ($p=0.292$). Poikien malli puolestaan on tilastollisesti merkitsevä ($p=0.026$) eli pojilla tämä muuttujayhdistelmä selittää merkitsevästi käännepesteen ajankohdan vaihtelua. Mallin yksittäisistä muuttujista käännepesteen ajankohtaa tilastollisesti merkitsevästi selittää ainoastaan äidin painoindeksi ennen raskautta (estimaatti -0.07 ja $p=0.001$). Pojilla äidin painoindeksin 1 kg/m^2 nousu aikaistaa käännepestettä merkitsevästi 0.07 vuodella eli vähän alle yhdellä kuukaudella. Mallin mukaan äidin painoindeksi oli siis yhteydessä aikaisempaan käännepesteseen. Tytöillä äidin painoindeksin 1 kg/m^2 nousu aikaistaa käännepestettä 0.04 vuodella eli noin puolella kuukaudella, mutta tämä yhteys ei ollut tilastollisesti merkitsevä. Kummassakaan mallissa diabetes ei ollut enää merkitsevä selittäjä, toisin kuin yhden selittäjän lineaarisissa regressiomalleissa. Diabetes-ryhmien väliset erot selittyvätkin ehkä suurilta osin äidin painoindeksillä, joka vaikuttaa erityisesti raskauden aikaisen diabeteksen syntyyn.

Toiseen useamman selittäjän lineaariseen regressiomallin lisättiin edellisten muuttujien lisäksi lapsen painoindeksin käännepesteen BMI. Mallit tehtiin taas erikseen tytöille ja pojille. Tyttöjen mallin tulokset on esitetty taulukossa 5.6 ja poikien malli tulokset taulukossa 5.7.

Sekä tyttöjen että poikien malli on tilastollisesti erittäin merkitsevä, poikien mal-

Taulukko 5.6. Tyttöjen useamman selittäjän lineaarisen regressiomallin tulokset. Mallissa selittäjänä mukana myös painoindeksin käännepisteen BMI. Mallin AIC=1236.2.

	Estimaatti	Keskivirhe	T-testisuure	P-arvo
Vakio	10.24	3.88	2.64	0.009
Äidin painoindeksi	-0.02	0.02	-0.86	0.391
Diabetes - Raskauden aikainen diabetes	-0.17	0.23	-0.76	0.449
Diabetes - Tyypin I diabetes	-0.80	0.72	-1.11	0.268
Äidin pituus	-0.01	0.01	-0.48	0.632
Syntymäpaino	-0.15	0.28	-0.54	0.591
Syntymäpituus	0.03	0.06	0.50	0.615
Päänympäryys	0.07	0.08	0.97	0.332
Piikin ajankohta	0.94	0.32	2.96	0.003
Piikin BMI	0.21	0.06	3.31	0.001
Raskaudenkesto (pv)	-0.01	0.01	-1.37	0.171
Äidin ikä (pv)	0.00	0.00	-1.13	0.261
Aiemmat synnytykset	0.03	0.09	0.29	0.775
Käännepisteen BMI	-0.48	0.08	-6.19	<0.001
Residuaalien keskivirhe	1.37			
Selitysaste	13.8 %			
Korjattu selitysaste	10.5 %			
F-testisuure	4.16			
P-arvo	<0.001			
Vapausasteet	338			

lissa selitysasteet tosin olivat jonkin verran parempia kuin tyttöjen mallissa. Molemmilla sukupuolilla mallien selitysasteet olivat käännepisteen BMI:n lisäyksen seurauksena yli tuplasti suurempia kuin edellisissä malleissa. Tyttöjen mallissa tilastollisesti merkitseviä selittäjiä olivat painoindeksin piikin ajankohta (estimaatti 0.94 ja $p=0.003$), painoindeksin piikin BMI (estimaatti 0.21 ja $p=0.001$) ja painoindeksin käännepisteen BMI (estimaatti -0.48 ja $p<0.001$). Painoindeksin käännepisteen BMI:n malliin lisäämisen seurauksena painoindeksin piikin ajankohdasta ja painoindeksin piikin BMI:stä tuli tilastollisesti merkitseviä. Poikien mallissa merkitseviä selittäjiä olivat äidin painoindeksi (estimaatti -0.05 ja $p=0.010$), painoindeksin piikin BMI (estimaatti 0.19 ja $p<0.001$) sekä käännepisteen painoindeksi (estimaatti -0.48 ja $p<0.001$). Painoindeksin piikin ajankohta ei pojilla ollut tilastollisesti merkitsevä, mutta painoindeksin piikin BMI:stä tuli tilastollisesti merkitsevä käännepisteen BMI:n malliin lisäämisen seurauksena. Mallien perusteella käännepisteen ajankohta näyttäisi siis olevan yhteydessä sitä edeltävään painoindeksin kehitykseen ja poikien kohdalla myös äidin painoindeksiin.

Mallien mukaan tytöillä painoindeksin piikin 1 kg/m²:n BMI:n nousu myöhäistää painoindeksin käännepistettä 0.21 vuodella ja painoindeksin käännepisteen 1 kg/m²:n BMI:n nousu aikaistaa käännepistettä 0.48 vuodella. Lisäksi painoindeksin piikin ajankohdan ollessa yhtä vuotta myöhäisempi, myös painoindeksin käännepiste ilmenee 0.94 vuotta myöhemmin. Pojilla painoindeksin käännepisteen 1 kg/m²:n

Taulukko 5.7. Poikien useamman selittäjän lineaarisen regressiomallin tulokset. Mallissa selittäjänä mukana myös painoindeksin käänne pisteen BMI. Mallin AIC=1399.8.

	Estimaatti	Keskivirhe	T-testisuure	P-arvo
Vakio	11.30	3.28	3.45	0.001
Äidin painoindeksi	-0.05	0.02	-2.59	0.010
Diabetes - Raskauden aikainen diabetes	-0.07	0.20	-0.36	0.720
Diabetes - Tyypin I diabetes	-0.52	0.56	-0.92	0.356
Äidin pituus	-0.01	0.01	-1.25	0.212
Syntymäpaino	-0.08	0.23	-0.33	0.741
Syntymäpituus	0.01	0.05	0.29	0.769
Pään ympärys	0.04	0.06	0.68	0.495
Piikin ajankohta	0.08	0.30	0.28	0.780
Piikin BMI	0.19	0.05	3.81	<0.001
Raskauden kesto (pv)	0.00	0.01	-0.01	0.994
Äidin ikä (pv)	0.00	0.00	-0.66	0.511
Aiemmat synnytykset	0.10	0.07	1.35	0.178
Käänne pisteen BMI	-0.48	0.07	-6.83	<0.001
Residuaalien keskivirhe	1.29			
Selitysaste	15.5 %			
Korjattu selitysaste	12.7 %			
F-testisuure	5.61			
P-arvo	<0.001			
Vapausasteet	399			

BMI:n nousu aikaistaa painoindeksin käänne pistettä 0.48 vuodella ja äidin painoindeksin 1 kg/m²:n nousu aikaistaa käänne pistettä 0.05 vuodella.

Painoindeksin piikin BMI:n parametrien estimaatit olivat positiivisia ja tilastollisesti merkitseviä molemmilla sukupuolilla eli tämän perusteella pienempi painoindeksi piikin aikaan edeltäisi aikaisempaa käänne pistettä. Vastaavanlainen löydös oli tehty myös joissakin aikaisemmissa tutkimuksissa (Eriksson et al. (2003) ja Eriksson et al. (2014)), joita tämä tutkimustulos tukee. Rolland-Cachera et al. (1984) ja Silverwood et al. (2009) puolestaan tekivät tutkimuksissaan hieman päinvastaiset löydökset. Painoindeksin käänne pisteen BMI:n parametrien estimaatit puolestaan olivat negatiivisia eli tämän perusteella suurempi painoindeksi käänne pisteen aikaan olisi yhteydessä aikaisempaan painoindeksin käänne pisteeseen. Tähän viittaavia tutkimustuloksia olikin saatu jo parissa aikaisemmassa tutkimuksessa (Dorosty et al. (2000) ja Eriksson et al. (2014)). Mielenkiintoista oli, että mallien perusteella ainoastaan pojilla oli havaittavissa äidin painoindeksin merkitsevä vaikutus käänne pisteen ajankohtaan. Lisäksi täytyy huomioda, että painoindeksin piikin BMI (ja tytöillä myös piikin ajankohta) yksistään ei ollut merkitsevä selittäjä käänne pisteen ajankohdalle, mutta yhdessä käänne pisteen BMI:n kanssa siitä tuli merkitsevä.

Lineaariset regressiomallit estimoitiiin käyttäen R:n "lm-funktiota. Mallien hyvyyteen liittyvä diagnostiikka on esitetty liitteessä C.

5.2.3 Lineaariset sekoitusregressiomallit

Muodostettiin seuraavaksi lineaarinen sekoitusregressiomalli. Malliin sisällytettiin vähemmän muuttujia kuin useamman selittäjän lineaarisiin regressiomalleihin. Selittäviä muuttujia jouduttiin jättämään pois, jotta saatiin aineisto jakautumaan komponentteihin. Useammalla selittävällä muuttujalla parhaaksi määräksi komponentteja valikoitui yksi eli tavallinen lineaarinen regressiomalli sopi parhaiten näihin analyyseihin. Sekoitusregressiomallia ei tehty erikseen tytöille ja pojille, jotta saatiin suurempi määrä havaintoja analysoitavaksi kerralla, mutta sukupuoli sisällytettiin malliin selittäväksi muuttujaksi. Malli sisälsi seuraavat selittävät muuttujat: äidin painoindeksi ennen raskautta, lapsen syntymäpainoindeksi, lapsen painoindeksin piikin ajankohta, lapsen painoindeksin piikin BMI sekä sukupuoli. Tällä muuttujayhdistelmällä parhaaksi malliksi valikoitui Akaiken informaatiokriteerin perusteella kaksi komponenttia sisältävä sekoitusregressiomalli. Sovitettiin kummallekin komponentille oma lineaarinen regressiomallinsa ja saatiin näin kaksi komponenttia sisältävä sekoitusregressiomalli. Sekoitusregressiomallin tulokset on esitetty taulukossa 5.8. Mallin parametrien estimaattien lisäksi taulukossa on esitetty, kuinka suuren osan aineistosta kumpikin komponentti sisältää sekä kumpaankin komponenttiin kuuluvien lasten painoindeksin keskimääräinen käännepisteen ajankohta. Lisäksi mallin parametrien estimaattien perään on merkattu niiden merkitsevyys ($. = p < 0.10$, $* = p < 0.05$, $*** = p < 0.001$). Taulukossa 5.9 on esitetty vertailun vuoksi yhden komponentin sekoitusregressiomallin eli tavallisen lineaarisen regressiomallin tulokset samoilla selittävillä muuttujilla. Lisäksi taulukossa 5.10 on esitetty sekoitusregressiomallin selittävien muuttujien keskiarvot sekä sukupuolten jakauma komponenteittain.

Taulukko 5.8. Kaksi komponenttia sisältävän sekoitusregressiomallin tulokset. Kummallekin komponentille on sovitettu oma lineaarinen regressiomallinsa. Mallin AIC=3262.6.

	Comp 1	Comp 2
Komponentin koko, N (%)	442 (47.8 %)	482 (52.2 %)
Keskimääräinen käännepisteen ajankohta	6.78	4.49
Estimaatit		
Vakio	6.86 ***	7.45 ***
Äidin painoindeksi	-0.05 ***	-0.06 ***
Syntymäpainoindeksi	-0.04	-0.08
Piikin ajankohta	0.61 .	-0.25
Piikin BMI	0.05	-0.01
Sukupuoli - Tyttö	-0.09	-0.27 *

Taulukosta 5.8 nähdään, että lapset ovat jakautuneet melko tasaisesti komponenttien kesken. Mallin ensimmäiseen komponenttiin kuuluu 47.8 % lapsista ja vastaavasti toiseen komponenttiin 52.2 % lapsista. Ensimmäiseen komponenttiin kuuluvilla lapsilla painoindeksin käännepiste on ollut keskimäärin huomattavasti myö-

Taulukko 5.9. Yhden komponentin sisältävän sekoitusregressiomallin eli tavallisen lineaarisen regressiomallin tulokset vertailtavaksi kaksi komponenttia sisältävään malliin. AIC=3266.0.

N	924
	Estimaatit
Vakio	7.11***
Äidin painoindeksi	-0.06***
Syntymäpainoindeksi	-0.04
Piikin ajankohta	0.19
Piikin BMI	0.02
Sukupuoli - Tyttö	-0.17

Taulukko 5.10. Sekoitusregressiomallin selittävien muuttujien keskiarvot ja sukupuolten jakaumat komponenteittain.

	Comp 1	Comp 2
Keskiarvot		
Äidin painoindeksi	23.49	23.57
Syntymäpainoindeksi	13.93	13.82
Piikin ajankohta	0.66	0.67
Piikin BMI	17.70	17.65
Osuudet		
Tytöt	45.5 %	46.1 %
Pojat	54.5 %	53.9 %

hemmin kuin toiseen komponenttiin kuuluvilla, joilla painoindeksin käännepesteen voidaan todeta olleen aikaistunut. Molemmissa komponenteissa äidin painoindeksi ennen raskautta oli erittäin merkitsevä selittäjä (parametrin estimaatit -0.05 ja -0.06). Toisessa komponentissa sukupuoli on merkitsevä selittäjä niin, että sukupuolen tyttö parametrin estimaatti on -0.27 verrattuna poikiin. Muita merkitseviä selittäjiä mallista ei löytynyt. Mielenkiintoista on, että painoindeksin piikin ajankohdan sekä piikin BMI:n parametrien estimaatit olivat erimerkkisiä eri komponenteissa. Ensimmäisessä komponentissa molempien selittävien muuttujien parametrien estimaatit olivat positiivisia, kun toisessa komponentissa ne puolestaan olivat molemmat negatiivisia. Tilastollista merkitsevyyttä näistä estimaateista ei kuitenkaan löytynyt, vaikkakin ensimmäisessä komponentissa painoindeksin piikin ajankohta oli tilastollisesti melkein merkitsevä. Taulukosta 5.10 käy ilmi, että kaikkien selittävien muuttujien keskimääräiset arvot olivat lähes yhtä suuria kummassakin komponentissa. Lisäksi kummassakin komponentissa poikia oli hieman enemmän kuin tyttöjä.

Mallin perusteella äidin suurempi painoindeksi aikaistaa käännepeistettä. Tämä pätee mallin molemmissa komponenteissa. Ensimmäisessä komponentissa äidin pai-

noindeksin yhden kg/m^2 :n nousu aikaistaa lapsen painoindeksin käännepestettä 0.05 vuotta. Toisessa komponentissa vastaava luku oli 0.06 vuotta. Ensimmäisessä komponentissa, johon kuuluvilla lapsilla painoindeksin käännepesteen ajankohta oli ollut keskimäärin myöhemmin kuin toiseen komponenttiin kuuluvilla, painoindeksin piikin myöhäisemmällä ajankohdalla oli myös painoindeksin käännepestettä myöhäistävä vaikutus. Yhden vuoden myöhäisempi painoindeksin piikki myöhäistää käännepestettä 0.61 vuotta. Vaikutus oli tilastollisesti melkein merkitsevä. Myös koko aineistolle sovitetussa mallissa (Taulukko 5.9) painoindeksin piikin ajankohdan sekä piikin BMI:n parametrien estimaatit olivat positiivisia, samoin kuin tässä sekoitusregressiomallin ensimmäisessä komponentissa. Lapsista löytyi kuitenkin joukko, jolle sovitetussa lineaarisessa regressiomallissa näiden muuttujien parametrien estimaatit olivat negatiivisia, niin kuin nähdään sekoitusregressiomallin toisen komponentin estimateista. Tällä joukolla oli myös keskimäärin aikaistunut käännepeste. Lisäksi tähän komponenttiin kuuluvilla tytöillä painoindeksin käännepeste oli tilastollisesti merkitsevästi aikaisemmin kuin pojilla. Sukupuolen tyttö käännepestettä aikaistava vaikutus oli 0.27 vuotta verrattuna poikaan. Taulukosta 5.9 nähdään, että koko aineistolle sovitetussa lineaarisessa regressiomallissa kaikista selittäjistä merkitsevä on vain äidin painoindeksi.

Sekoitusregressiomallissa aineisto jakautui kahteen komponenttiin niin, että ensimmäisen komponentin lasten painoindeksin käännepeste oli selvästi myöhäisempi kuin toisen komponentin lasten, ja painoindeksin piikkiin liittyvien muuttujien estimaatit olivat erimerkkisiä näissä komponenteissa. Tämä antaa viitteitä siitä, että painoindeksin piikin ajankohdalla sekä piikin BMI:llä saattaa olla erilainen yhteys käännepesteen ajankohtaan riippuen tuntemattomasta taustatekijästä, joka jakaa aineiston näihin kahteen komponenttiin.

Lineaariset sekoitusregressiomallit estimoitiiin käyttäen R:n funktioita "stepFlexmix" ja "flexmix". "stepFlexmix"-funktion avulla selvitettiin parhaiten sopiva komponenttien määrä kyseisellä muuttujajyhdistelmällä. Funktio estimoi mallit määritellyille määrille komponentteja, tässä tapauksessa 1-10 komponentille ($k=1:5$), kullekin komponenttien lukumäärälle kymmeneen kertaan ($nrep=10$). Funktio antaa kullekin komponenttien lukumäärälle parhaan estimoidun mallin hyvyyteen liittyvät informaatiokriteeriarvot (AIC, BIC ja ICL), joiden perusteella määritettiin paras komponenttien lukumäärä. Käyttäen "flexmix"-funktiota muodostettiin lineaarinen sekoitusregressiomalli valitulla komponenttien lukumäärällä. Mallin parametrien estimaattien tilastolliset merkitsevyydet saatiin "refit"-funktion avulla. Mallin hyvyyteen liittyvä diagnostiikka eli posterioritodennäköisyyksien rootogrammi on esitetty liitteessä C.

6 Johtopäätelmät

Tutkielman tarkoituksena oli mallintaa lasten painoindeksejä iän suhteen siihen sopivalla tavalla sekä estimoida mallien avulla jokaiselle lapselle painoindeksin varhaislapsuuden piikki sekä painoindeksin käännepiste. Mallinnus päätettiin tehdä kahdessa osassa käyttäen polynomiregressiomalleja sekä kuutiotasoitettuja splineja. Polynomiregressiomallien avulla estimoitiin painoindeksin varhaislapsuuden piikit ja kuutiotasoitettujen splinien avulla estimoitiin painoindeksin käännepisteet kullekin lapselle.

Toisena tarkoituksena oli tutkia painoindeksin käännepisteen ajankohtaan yhteydessä olevia tekijöitä. Tutkittaviksi tekijöiksi valittiin saatavilla olevista tekijöistä sellaiset, joilla voisi mahdollisesti olla yhteys lapsen käännepisteen ajankohtaan aikaisemman kirjallisuuden perusteella. Tekijät olivat siis joko sellaisia, joilla oli jo todettu olevan jonkinlainen yhteys käännepisteeseen tai löydettyjen yhteyksien pohjalta muuten valittuja. Aiheesta löytyy tutkimuksia vielä sen verran vähän, että on mielekästä tutkia myös jo aikaisemmin tutkittuja asioista, koska näitä löydöksiä ei ole vielä monissa tutkimuksissa toistettu. Lisäksi aikaisemmissa tutkimuksissa oli löydetty joitakin hieman ristiriitaisia tuloksia.

Aineistossa olisi saattanut olla myös muutamia muita lapsen käännepisteen ajankohdan kannalta kiinnostavia muuttujia, mutta osan kohdalla ongelmaksi muodostui puuttuvan tiedon määrä. Tällaisia olivat esimerkiksi äidin imetys sekä äidin loppuraskauden painoindeksi tai tämän perusteella laskettu äidin painonnousu raskauden aikana. Äidin loppuraskauden painoindeksin yhteyttä käännepisteen ajankohtaan tutkittiin silti yhden muuttujan lineaarisen regressiomallin avulla, mutta tällöin alkuperäisestä aineistosta oli käytössä vain alle puolet.

Vaikka tutkielman malleista esille nousi muutamia käännepisteen ajankohtaa merkitsevästi selittäviä muuttujia, mallien selitysasteet eivät olleet kovin suuria. Tämä viittaa siihen, että tutkimuksen kohteena olleet käännepistettä mahdollisesti selittävät tekijät eivät edes yhdessä selitä kovin suurta osaa lapsen käännepisteen ajankohdan vaihtelusta. Tutkielmassa tosin tarkasteltiin vain muuttujien välisiä lineaarisia riippuvuuksia, joten ei voida varmaksi todeta, ettei niiden välille olisi löydetävissä selkeämpiä epälineaarisia yhteyksiä.

Tutkimuksen perusteella äidin painoindeksi näyttäisi vaikuttavan lapsen painoindeksin ajankohtaan niin, että äidin suurempi painoindeksi altistaa lapsen aikaisemmalle käännepisteen ajankohdalle. Tämä olikin hyvin odotettu tulos, joka oli todettu myös aikaisemmissa tutkimuksissa. Äidin painoindeksin yhteys käännepisteen ajankohtaan oli erityisesti nähtävissä pojilla. Mielenkiintoista oli, että tytöillä yhteys ei ollut tilastollisesti merkitsevä, kun malliin lisättiin useampi selittävä muuttuja. Lisäksi äidin diabetekselle ja painoindeksin käännepisteen ajankohdalle löytyi yhteys, joka kuitenkin näytti katoavan, kun malliin lisättiin äidin painoindeksi. Tämä luultavasti johtui siitä, että todellinen selittävä tekijä on äidin painoindeksi, joka puolestaan vaikuttaa vahvasti raskauden aikaisen diabeteksen puhkeamiseen.

Tutkimus antoi myös viitteitä siitä, että lapsen painoindeksin kehityksen vaiheet

vaikuttavat painoindeksin käännepesteen ajankohtaan. Sekä tytöillä että pojilla painoindeksin varhaislapsuuden piikin BMI:n sekä painoindeksin käännepesteen BMI:n ollessa samassa mallissa, ne molemmat olivat tilastollisesti merkitsevästi yhteydessä käännepesteen ajankohtaan. Molemmilla sukupuolilla painoindeksin piikin aikainen alhaisempi BMI sekä painoindeksin käännepesteen aikainen korkeampi BMI altistivat lapsen aikaisemmalle painoindeksin käännepesteele. Lisäksi tytöillä painoindeksin piikin aikaisempi ajankohta altisti lapsen aikaisemmalle käännepesteele. Pojillakin painoindeksin piikin ajankohdan vaikutus oli samansuuntainen, mutta ei merkitsevä. Näiden perusteella voisi olla mahdollista, että rasvasolujen kasvun loppuessa aikaisemmin lapsen ensimmäisen elinvuoden aikana, ne alkaisivat myös kasvaa uudelleen aikaisemmin eli käännepeste ilmentyisi tämän seurauksena aikaisemmin. On kuitenkin hyvä pitää mielessä, että painoindeksin piikin ajankohta tai piikin BMI eivät olleet tilastollisesti merkitsevästi yhteydessä painoindeksin käännepesteen ajankohtaan niiden ollessa malleissa ilman painoindeksin käännepesteen BMI:ä. Yksistään ne eivät siis selittäneet aikaistunutta käännepestettä. Yhteys käännepesteen suuremman BMI:n ja aikaisemman käännepesteen ajankohdan välillä voisi viitata siihen, että painoindeksin laskun loppuminen varhaislapsuuden piikin jälkeen johtaisi käännepesteen ilmentymiseen. Painoindeksin laskun loppuessa aikaisemmin, käännepesteen aikainen painoindeksi jäisi suuremmaksi.

Sekoitusregressiomallin avulla saatiin myös mielenkiintoisia tuloksia. Malli jakoi aineiston kahteen komponenttiin, joista ensimmäiseen kuuluvien lasten painoindeksin käännepeste oli keskimäärin ollut yli kaksi vuotta myöhemmin kuin toiseen komponenttiin kuuluvien lasten. Mielenkiintoista oli, että painoindeksin piikin ajankohdan sekä piikin BMI:n vaikutus käännepesteen ajankohtaan oli erisuuntaista näissä komponenteissa. Ensimmäisen komponentin lapsilla aikaisempi painoindeksin piikki sekä pienempi piikin BMI ennakoivat aikaisempaa käännepestettä kun taas toisen komponentin lapsilla aikasempaa käännepestettä ennakoiva kasvumalli oli päinvastainen. Yhteydet eivät kuitenkaan olleet tilastollisesti merkitseviä, joten niistä ei voida suurempia johtopäätöksiä tehdä. Komponenttien erot käännepesteen keskimääräisessä ajankohdassa saattavat kuitenkin viitata siihen, että taustalla voisi olla jokin tuntematon tekijä, joka jakoi aineiston näihin kahteen tiettyyn komponenttiin, joissa aikaisempaan käännepesteeeseen johtavat kasvumallit olivat keskenään erilaiset.

Tässä tutkimuksessa ei löydetty viitteitä lapsen syntymämittausten yhteydestä painoindeksin käännepesteen ajankohtaan. Lapsen painon ja painoindeksin kehitysvaiheita olisi syytä tutkia vielä lisää, koska selvää on, että lihavuuden kehitys alkaa jo hyvin varhaisessa vaiheessa. On kuitenkin otettava huomioon, että geneettiset tekijät ovat myös jonkinlaisessa osassa lihavuuden kehittymisessä ja niihin emme pysty itse vaikuttamaan. Taylor ym. (2005) toteavatkin painoindeksin käännepesteeeseen liittyvässä kirjallisuuskatsauksessaan, että koska ihmisen kasvun ja kehityksen kaavat ovat vahvasti riippuvia geeneistä, on mahdollista, että myös käännepesteen ajankohta on pitkälti riippuvainen geneettisistä tekijöistä. Ainakin Sovio, Mook-Kanamori, Warrington, Lawrence, Briollais, Palmer, Cecil, Sandling, Syvänen, Kaakinen, Beilin, Millwood, Bennett, Laitinen, Pouta, Molitor, Davey Smith, Ben-Shlomo, Jaddoe, Palmer, Pennell, Cole, McCarthy, Järvelin, Timpson & Con-

sortium (2011) ovat julkaisseetkin jo aiheesta artikkelin. Myös geenien yhteyttä käännepisteeseen olisi mielenkiintoista tutkia tulevaisuudessa lisää.

Tutkielman lopuksi haluan esittää kiitokset erityisesti Tiina Hakaselle, jonka kautta sain Pro gradu -tutkielmani aiheen sekä ohjaajalleni Tapio Nummelle, josta oli suuri apu tilastollisten menetelmien valinnassa ja toteuttamisessa. Kiitos myös Jaakko Peltoselle säännöllisten graduseminaaritapaamisten järjestämisestä.

Lähteet

- Dorosty, A. R., M. Emmett, P., Cowin, I. S. & Reilly, J. J. (2000), "Factors Associated With Early Adiposity Rebound.", *Pediatrics*, 105, 1115–1118.
- Efromovich, S. (1999), *Nonparametric Curve Estimation*, Secaucus, US: Springer.
- Eriksson, J. G., Forsén, T., Tuomilehto, J., Osmond, C. & Barker, D. J. P. (2003), "Early adiposity rebound in childhood and risk of Type 2 diabetes in adult life", *Diabetologia*, 46, 190–194. Saatavilla Internetistä: <http://dx.doi.org/10.1007/s00125-002-1012-5>.
- Eriksson, J. G., Kajantie, E., Lampl, M., Osmond, C. & Barker, D. J. P. (2014), "Small head circumference at birth and early age at adiposity rebound.", *Acta Physiologica*, 210, 154–160.
- Harjunmaa, U. (2009), "Current growth patterns of Finnish children aged 0 to 4 years", pro gradu -tutkielma, Lääketieteen yksikkö, Tampereen yliopisto. Saatavilla Internetistä: <https://uta32-kk.lib.helsinki.fi/bitstream/handle/10024/81098/gradu03930.pdf?sequence=1>.
- Hof, M. H. P., Vrijkotte, T. G. M., de Hoog, M. L. A., van Eijnden, M. & Zwinderman, A. H. (2013), "Association between Infancy BMI Peak and Body Composition and Blood Pressure at Age 5?6 Years", *PLoS ONE*, 8, 1–11. Saatavilla Internetistä: <http://dx.doi.org/10.1371/journal.pone.0080517>.
- Hughes, A. R., Sherriff, A., Ness, A. R. & Reilly, J. J. (2014), "Timing of Adiposity Rebound and Adiposity in Adolescence", *Pediatrics*, 134, e1354–e1361. Saatavilla Internetistä: <http://pediatrics.aappublications.org/content/134/5/e1354>.
- Johnson, W., Soloway, L. E., Erickson, D., Choh, A. C., Lee, M., Chumlea, W. C., Siervogel, R. M., Czerwinski, S. A., Towne, B. & Demerath, E. W. (2012), "A changing pattern of childhood BMI growth during the 20th century: 70 y of data from the Fels Longitudinal Study", *The American Journal of Clinical Nutrition*, 95, 1136–1143. Saatavilla Internetistä: <http://ajcn.nutrition.org/content/95/5/1136.abstract>.
- Koyama, S., Kariya, K., Ichikawa, G., Shimura, N., Sairenchi, T. & Arisaka, O. (2013), "Factors affecting the timing of adiposity rebound", *International Journal of Pediatric Endocrinology*, 2013, 1–1. Saatavilla Internetistä: <http://dx.doi.org/10.1186/1687-9856-2013-S1-O47>.
- Leisch, F. (2004), "FlexMix: A General Framework for Finite Mixture Models and Latent Class Regression in R", 11. Saatavilla Internetistä: <https://www.jstatsoft.org/article/view/v01i08l>.
- McLachlan, G. & Peel, D. (2004), *Finite Mixture Models* (Wiley Series in Probability and Statistics), Wiley.
- Männistö, S., Laatikainen, T. & Vartiainen, E. (2012), "Suomalaisten lihavuus ennen ja nyt", Terveiden ja hyvinvoinninlaitos. Saatavilla Internetistä: http://www.julkari.fi/bitstream/handle/10024/90885/TutkimuksestaTiiviisti4_lihavuus.pdf?sequence=1.
- Montgomery, D. C., Peck, E. A. & Vining, G. G. (2006), *Introduction to Linear Regression Analysis* (Wiley Series in Probability and Statistics), Wiley.
- Mook-Kanamori, D. O., Durmus, B., Sovio, U., Hofman, A., Raat, H., Steegers, E. A., Järvelin, M. R. & Jaddoe, V. W. (2011), "Fetal and infant growth and the risk of obesity

- during early childhood: the Generation R Study.”, *European journal of endocrinology / European Federation of Endocrine Societies*, 165, 623–630. Saatavilla Internetistä: <http://www.eje-online.org/content/165/4/623.full>.
- Mustajoki, P. (2015), ”Painoindeksi (BMI)”, teoksessa *Terveyskirjasto*. Saatavilla Internetistä: http://www.terveyskirjasto.fi/terveyskirjasto/tk.koti?p_artikkeli=dlk01001.
- (NCD-RisC), N. R. F. C. (2016), ”Trends in adult body-mass index in 200 countries from 1975 to 2014: a pooled analysis of 1698 population-based measurement studies with 19.2 million participants”, *The Lancet*, 387, 1377 – 1396. Saatavilla Internetistä: <http://www.sciencedirect.com/science/article/pii/S014067361630054X>.
- Nummi, T., Hakanen, T., Lipiäinen, L., Harjunmaa, U., Salo, M. K., Saha, M.-T. & Vuorela, N. (2014), ”A trajectory analysis of body mass index for Finnish children”, *Journal of Applied Statistics*, 41, 1422–1435.
- Rolland-Cachera, M. F., Deheeger, M., Akrou, M. & Bellisle, F. (1995), ”Influence of macronutrients on adiposity development: a follow up study of nutrition and growth from 10 months to 8 years of age”, *International Journal of Obesity and Related Metabolic Disorders : Journal of the International Association for the Study of Obesity*, 19, 573–578.
- Rolland-Cachera, M. F., Deheeger, M., Bellisle, F., Bellisle, F., Sempe, M., Guillaud-Bataille, M. & Patois, E. (1984), ”Adiposity rebound in children: a simple indicator for predicting obesity”, *The American Journal of Clinical Nutrition*, 39, 129–135.
- Rolland-Cachera, M. F., Deheeger, M., Mailliot, M. & Bellisle, F. (2006), ”Early adiposity rebound: causes and consequences for obesity in children and adults”, *International Journal of Obesity*, 30, 11–17. Saatavilla Internetistä: <http://www.nature.com/ijo/journal/v30/n4s/full/0803514a.html>.
- Silverwood, R. J., De Stavola, B. L., Cole, T. J. & Leon, D. A. (2009), ”BMI peak in infancy as a predictor for later BMI in the Uppsala Family Study”, *International Journal of Obesity*, 33, 929–937. Saatavilla Internetistä: <http://dx.doi.org/10.1038/ijo.2009.108>.
- Simonoff, J. (2012), *Smoothing Methods in Statistics* (Springer Series in Statistics), Springer New York.
- Sovio, U., Mook-Kanamori, D. O., Warrington, N. M., Lawrence, R., Briollais, L., Palmer, C. N. A., Cecil, J., Sandling, J. K., Syvänen, A.-C., Kaakinen, M., Beilin, L. J., Millwood, I. Y., Bennett, A. J., Laitinen, J., Pouta, A., Molitor, J., Davey Smith, G., Ben-Shlomo, Y., Jaddoe, V. W. V., Palmer, L. J., Pennell, C. E., Cole, T. J., McCarthy, M. I., Jarvelin, M.-R., Timpson, N. J. & Consortium, E. G. G. (2011), ”Association between Common Variation at the Locus and Changes in Body Mass Index from Infancy to Late Childhood: The Complex Nature of Genetic Association through Growth and Development”, *PLoS Genet*, 7, 1–13. Saatavilla Internetistä: <http://dx.doi.org/10.1371/journal.pgen.1001307>.
- Taylor, R. W., Grant, A. M., Goulding, A. & Williams, S. M. (2005), ”Early adiposity rebound: review of papers linking this to subsequent obesity in children and adults”, *Current Opinion in Clinical Nutrition and Metabolic Care*, 8, 607–612.
- Vignerová, J., Humeníková, L., Brabec, M., Riedlová, J. & Bláha, P. (2007), ”Long-term changes in body weight, BMI, and adiposity rebound among children and adolescents in the Czech republic”, *Economics and Human Biology*, 5, 409–425. Special Issue on Obesity in Eastern Europe.

- Weisberg, S. (2013), *Applied Linear Regression* (Wiley Series in Probability and Statistics), Wiley.
- Wen, X., Kleinman, K., Gillman, M. W., Rifas-Shiman, S. L. & Taveras, E. M. (2012), "Childhood body mass index trajectories: modeling, characterizing, pairwise correlations and socio-demographic predictors of trajectory characteristics", *BMC Medical Research Methodology*, 12, 1–13. Saatavilla Internetistä: <http://bmcmmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-12-38>.
- Whitaker, R. C., Pepe, M. S., Wright, J. A., Seidel, K. D. & Dietz, W. H. (1998), "Early Adiposity Rebound and the Risk of Adult Obesity", *Pediatrics*, 101, e5. Saatavilla Internetistä: <http://pediatrics.aappublications.org/content/101/3/e5>.
- Wood, S. (2006), *Generalized Additive Models: An Introduction with R* (Chapman & Hall/CRC Texts in Statistical Science), Taylor & Francis.

Liite A

Muuttujalista

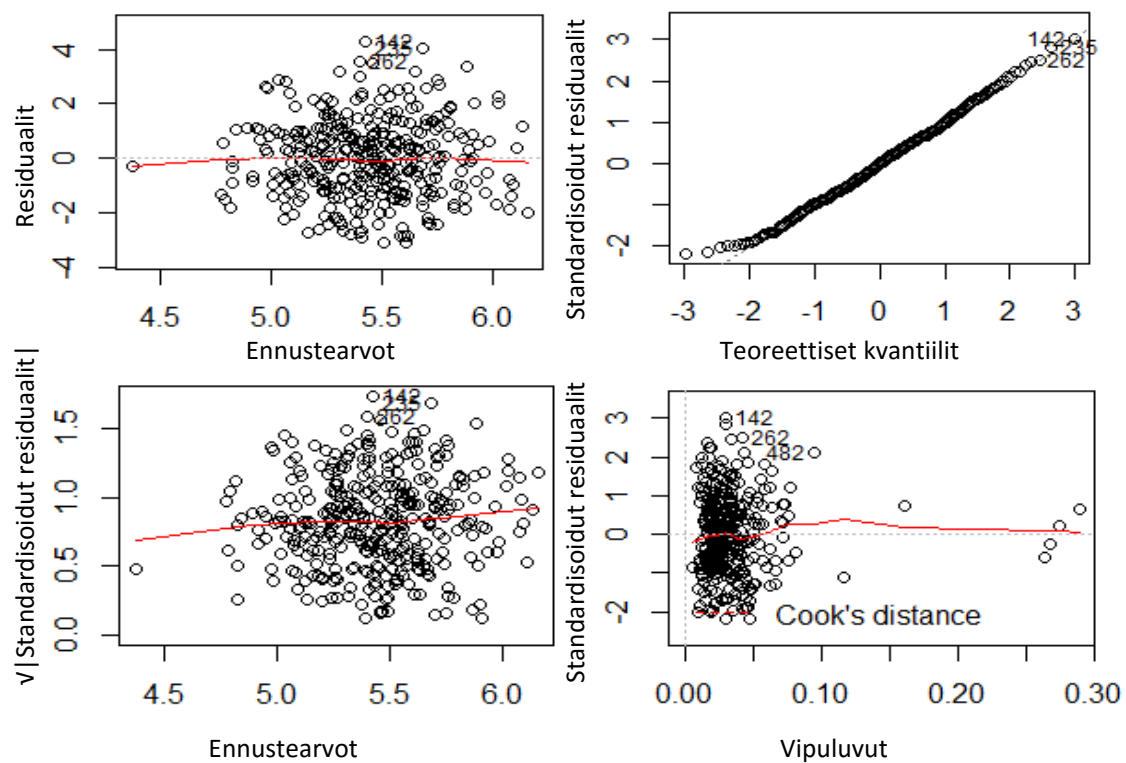
Muuttuja	Selite	Arvot
child	Lapsen tunnus	
vuosi	Lapsen syntymävuosi	
sukupuoli	Lapsen sukupuoli	Poika; Tyttö;
syntymapaino	Lapsen syntymäpaino	
syntymäpituus	Lapsen syntymäpituus	
syntymaBMI	Lapsen syntymäpainoindeksi	
paanympäry	Lapsen syntymäpäänympäry	
aiemmatraskaudet	Äidin aiempien raskauksien lkm	
keskenmenot	Äidin keskenmenojen lkm	
keskeytykset	Äidin raskaudenkeskeytysten lkm	
ulkopuolisetraskaudet	Äidin kohdun ulkopuolisten raskauksien lkm	
aiemmat synnytykset	Äidin aiempien synnytysten lkm	
kuolleenasyntyneet	Äidin kuolleenä syntyneiden lasten lkm	
aiti_ika	Äidin ikä synnyttäessä (vuosina)	
aiti_ika_pv	Äidin ikä synnyttäessä (päivinä)	
diabetes	Äidin diabetes	Ei; Raskauden aikainen diabetes; Tyypin I diabetes;
tupakointi	Äidin tupakointi raskauden aikana	Ei; Lopetti 1. kolmanneksen aikana; Tupakoinut 1. kolmanneksen jälkeen;
raskaudenkesto_pv	Raskauden kesto päivinä	
imetys_synnyty	Imetys synnytyksen jälkeen sairaalasta lähdettäessä	Ei; Kyllä; Osittain;
imetys_jalkitark	Imetys jälkitarkastuksessa	Ei; Kyllä; Osittain;
aidinpaino	Äidin paino ennen raskautta	
aidinpituus	Äidin pituus	
aidinBMI_er	Äidin painoindeksi ennen raskautta	Alipaino;
aidinBMI_er_1k6	Äidin luokiteltu painoindeksi ennen raskautta	Normaali paino; Ylipaino; Merkittävä lihavuus; Vaikea lihavuus; Sairaalloinen lihavuus;
aidinBMI_er_1k4	Äidin luokiteltu painoindeksi ennen raskautta	Alipaino; Normaali paino; Ylipaino; Lihavuus;

aidinBMI_er_lk3	Äidin luokiteltu painoindeksi ennen raskautta	Alipaino; Normaali paino; Ylipaino tai lihavuus;
aidin_paino_viimeinen	Äidin paino viimeisellä raskaudenseurantakäynnillä	
aidin_BMI_viimeinen	Äidin painoindeksi viimeisellä raskaudenseurantakäynnillä	
rv_viimBMImittaus_pv	Raskauden kesto päivinä viimeisellä raskaudenseurantakäynnillä	
painonmuutos	Äidin painon muutos raskauden aikana	
BIMmuutos	Äidin painoindeksin muutos raskauden aikana	
Estimoidut muuttujat		
ARtime	Lapsen ikä vuosina painoindeksin käännepisteessä	
ARbmi	Lapsen BMI painoindeksin käännepisteessä	
IPtime	Lapsen ikä vuosina painoindeksin varhaislapsuuden piikissä	
IPbmi	Lapsen BMI painoindeksin varhaislapsuuden piikissä	

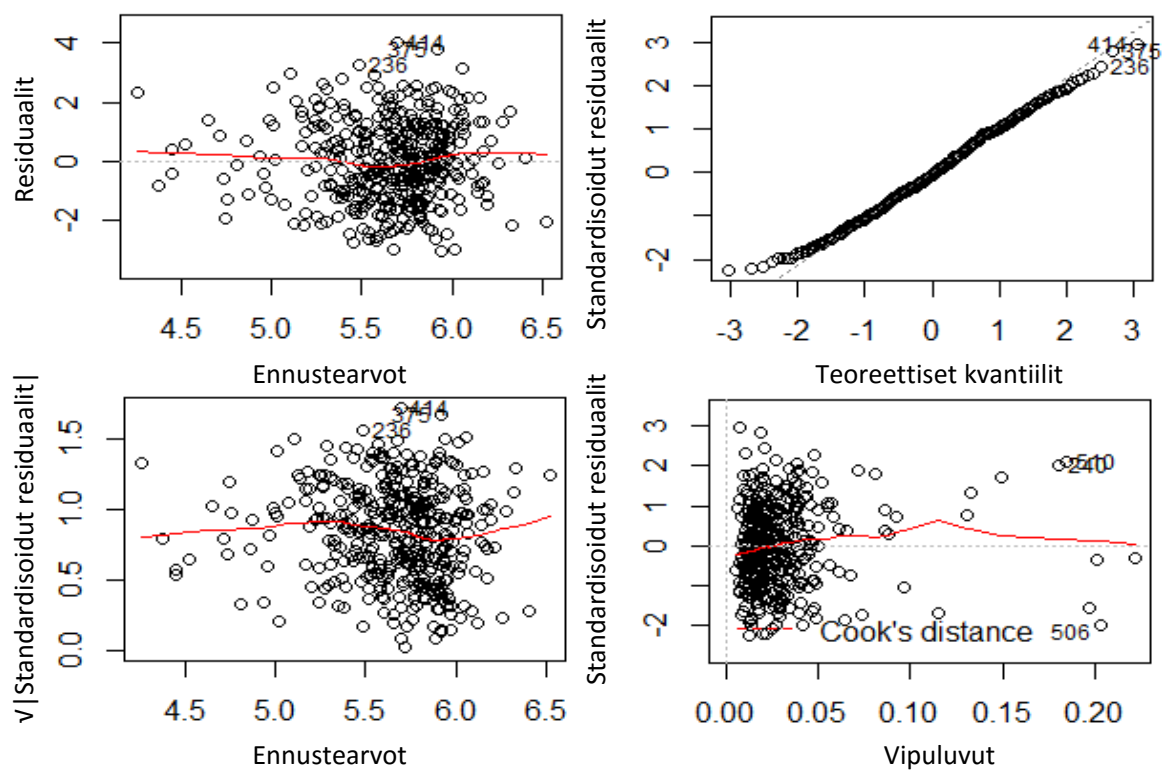
Liite C

Mallien diagnostiikkaa

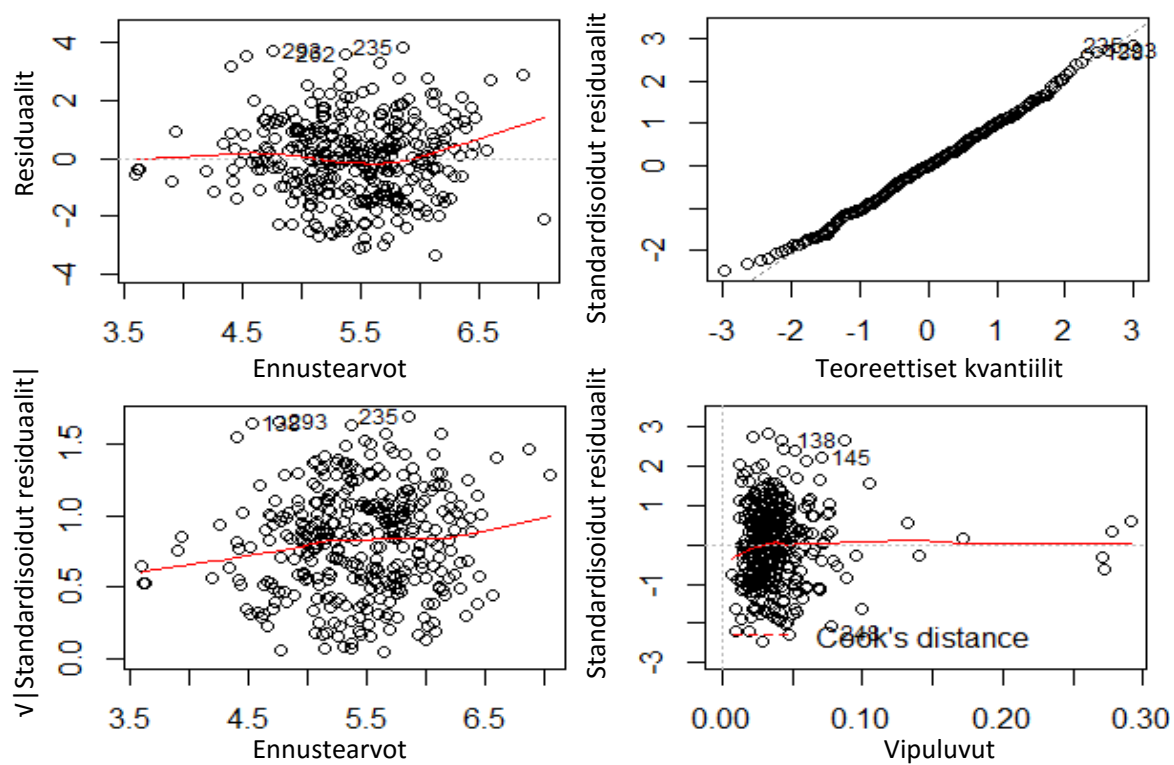
C.1 Malli 5.4



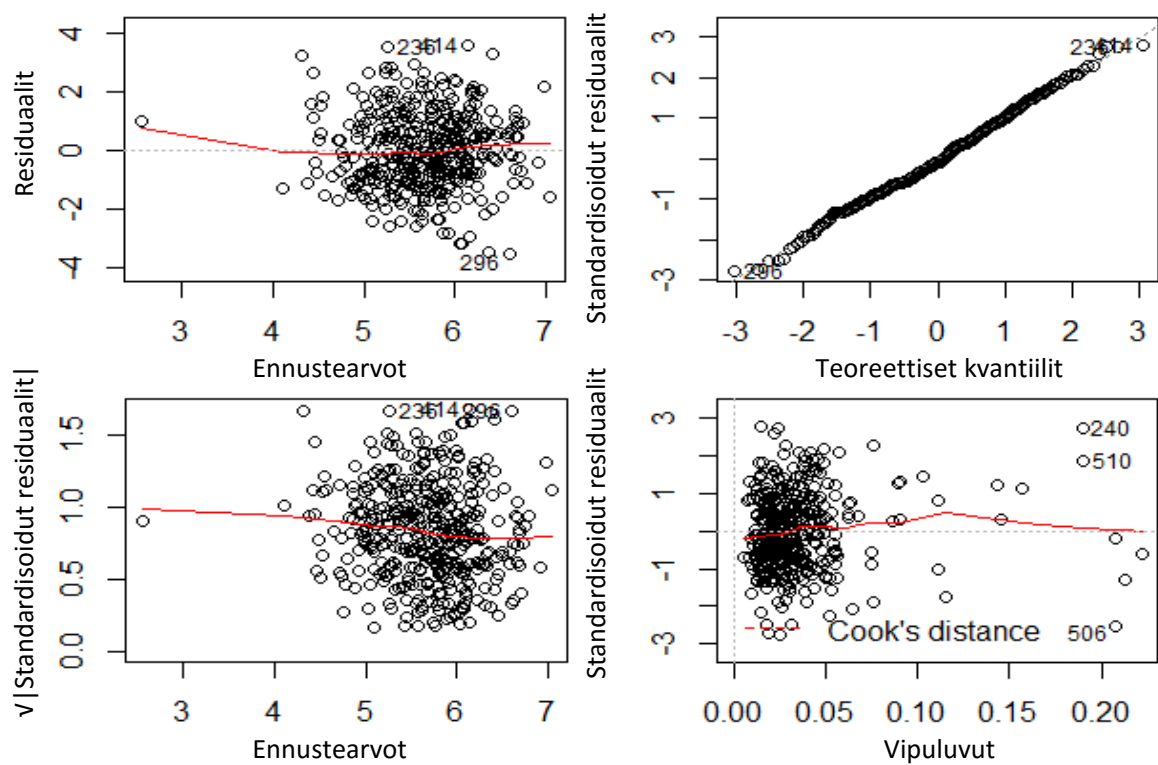
C.2 Malli 5.5



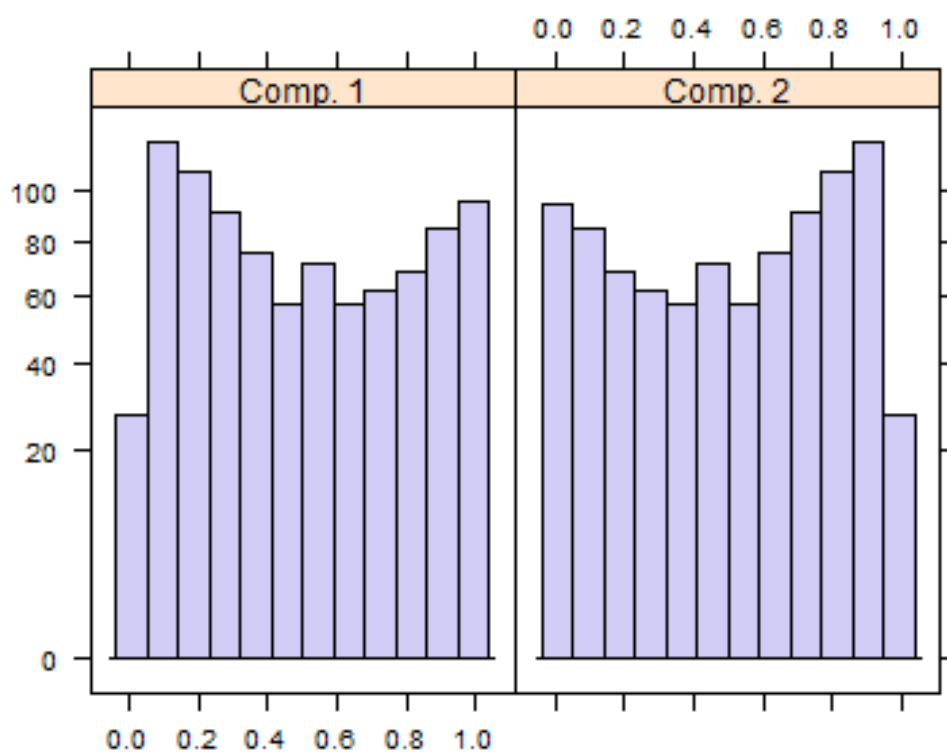
C.3 Malli 5.6



C.4 Malli 5.7



C.5 Sekoitusregressiomalli



Posterioritodennäköisyyksien ($>1e-04$) rootogrammi